BORRELIDIN-PRODUCING POLYKETIDE SYNTHASE AND ITS USES

FIELD OF THE INVENTION

The present invention relates to materials and methods for the preparation of polyketides. Enzyme systems, nucleic acids, vectors and cells are provided for the preparation of polyketides, and in particular the polyketide macrolide borrelidin.

BACKGROUND TO THE INVENTION

Polyketides are natural products produced by a wide range of organisms, and particularly by microorganisms. Polyketides have many important pharmaceutical, veterinary and agricultural uses. Polyketides encompass a huge range of chemical structural space, and have a wide range of associated biological activities. Polyketides with use in medical treatments include antibiotics, immunosuppressants, antitumor agents, other chemotherapeutic agents, and other compounds possessing a broad range of therapeutic and biological properties. The Gram-positive bacteria *Streptomyces* and their allied genera are prodigious producers of polyketides, and the genetics and biochemistry of polyketide biosynthesis in these organisms are relatively well characterised (Hopwood, 1997). The genes for polyketide biosynthesis in *Streptomyces* are clustered and the exploitation of DNA technology has made it possible to isolate complete biosynthetic gene clusters by screening gene libraries with DNA probes encoding the genes responsible for their biosynthesis. Thus, increasing numbers of gene clusters for polyketide biosynthesis in *Streptomyces* and other microorganisms have been isolated and sequenced, including, for example, those for the polyether monensin (WO 01/68867), the polyene nystatin (WO 01/59126) and for rapamycin (Schwecke *et al.*, 1995).

Polyketides are synthesised through the repeated condensation of building blocks that contain a carboxylic acid function. At each stage of the process this results in the formation of a new β-keto function and an α-side chain branch into the growing chain. The structural diversity of polyketides derives from a number of aspects of their biosynthetic pathway including: the wide variety of starter units that may be utilised in their biosynthesis; the different lengths of polyketide chains that are possible; the various α-side chains that are introduced either during or after assembly of the polyketide chain; the various β-substitutions that may be introduced during or after assembly of the polyketide chain; the various degrees of processing that the β -keto

groups can undergo (keto, hydroxyl, enoyl, and methylene); and the various stereochemistries that are possible at the α- and β -centres.

The synthesis of polyketides is catalysed by an enzyme, or by a complex of enzymes, called the polyketide synthase (PKS) in a manner similar to that of fatty acid biosynthesis. *Streptomyces* and related genera PKSs fall into three main categories: type-I, type-II and type-III. The type-III PKSs are small proteins related to plant chalcone synthases that have been discovered only recently (Moore & Hopke, 2000). Type-III systems have been implicated in the biosynthesis of a small number of secondary metabolites but may be more generally involved in the biosynthesis of soluble pigments (Cortés et al., 2002). The type-II PKSs consist of several monofunctional proteins that act as a multi-polypeptide complex. Simple aromatic polyketides such as actinorhodin are formed by several rounds of chain assembly, which are performed iteratively on one set of type-II PKS enzymes that are encoded for by one set of PKS genes (Hopwood, 1997). Type-I PKSs are multifunctional proteins and are required for the synthesis of more complex polyketides such as erythromycin and rapamycin. As the focus of this patent, type-I PKS organisation and function are described in detail below:

Type-I PKSs are organised into modules, whereby each module consists of several catalytic 'domains' that are required to carry out one round of chain assembly (Staunton & Wilkinson, 1997). In general a modular PKS contains the correct number of modules (loading plus extension modules) to select and condense the correct number of loading and extension units. For example the erythromycin PKS consists of 7 modules (one loading and six extension modules) to select and condense the one starter and six extension units required for the biosynthesis of the erythromycin precursor 6-deoxyerythronolide B. Thus, there exists a one to one relationship between the number of modules present in the PKS and the number of units incorporated. This one to one relationship is described as 'co-linearity'.

The term 'extension module' as used herein refers to the set of contiguous domains, from the β-ketoacyl-acyl carrier protein synthase (KS) domain to the next acyl carrier protein (ACP) domain, which accomplishes one cycle of polyketide chain extension. The term 'loading module' as used herein refers to any group of contiguous domains that accomplishes the loading of the starter unit onto the PKS and thus renders it available to the KS domain of the first extension module. Besides condensation of the next extender carboxylic acid (or ketide) unit onto the growing

polyketide chain, which is performed by the catalytic activity of the essential KS domain, modules of type-I PKSs may contain domains with β-ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) activities which are responsible for the further processing of the newly formed β-keto groups during chain extension. The acyl transferase (AT) and the ACP domains present in each module are responsible for the choice of extender unit, and the tethering of the growing chain during its passage on the PKS respectively. The AT domains of a modular PKS can also be found as discrete proteins (Cheng et al., 2003). The completed polyketide chain is generally released from PKSs by the action of a terminal thioesterase (TE) domain that is also generally involved in the cyclisation (lactonisation) of the final product. Other chain terminating/cyclising strategies are also employed such as that for the addition of an amino acid residue and macrolactam formation as observed for rapamycin (Schwecke *et al.*, 1995), for macrolactam formation as for rifamycin (August *et al.*, 1998), and for amino acid incorporation followed by reductive elimination as for myxalamid biosynthesis (Silakowski *et al.*, 2001). In summary, there is a single enzymatic domain present for each successive catalytic step that occurs during biosynthesis on the PKS, and they are used in defined sequence that depends upon their location within the protein and the particular function they perform. This mechanism is termed 'processive'.

The modular arrangement of type-I PKSs was first confirmed by mutation of the erythromycin PKS (also known as 6-deoxyerythronolide B synthase, DEBS) through an in-frame deletion of a region of the KR domain of module 5 (Donadio *et al.*, 1991). This led to the production of the erythromycin analogues, 5,6-dideoxy-3-α-mycarosyl-5-oxoerythronolide B and 5,6-dideoxy-5-oxoerythronolide B, due to the inability of the mutated KR domain to reduce the β-keto group 5 at this stage of processive biosynthesis. Likewise, alteration of the active site residues in the ER domain of module 4 of DEBS2, by genetic engineering of the corresponding PKS-encoding DNA and its introduction into *Saccharopolyspora erythraea*, led to the production of 6,7-anhydroerythromycin C (Donadio *et al.*, 1993). In addition, the length of the polyketide chain formed by DEBS has been altered through the specific relocation of the TE domain of DEBS3 to the end of DEBS1; the expected triketide lactone product was produced in good yield (Cortés *et al.*, 1995). It should be noted that the changes described involved modification by deletion of sequence, or by sequence specific inactivation, or by the alternative juxtaposition of DNA sequence from within the same

PKS cluster (ie. they are considered '*homologous changes*'). Other such 'homologous' changes to the erythromycin PKS are described in WO 93/13663.

The modular organisation of type-I PKS genes lends itself to the manipulation of these genes to produce altered polyketide structures. Type I PKSs represent an assembly line for polyketide biosynthesis that can be manipulated by changing the number of modules; by changing their specificities towards different carboxylic acid starter units and extender units; by inactivating, mutating, removing, swapping or inserting domains with different activities and specificities; and by altering the chain or ring size through the repositioning of termination or cyclisation domains (Staunton & Wilkinson, 1997).

WO 98/01546 describes the production of hybrid PKS gene assemblies comprising the incorporation of *heterologous* DNA. WO 98/01546 describes methods for generating hybrid PKSs in which the substitution of genes encoding heterologous modules, sub-modules or domains for the native genes generates novel polyketides with altered structures. Specifically, for example the AT domains of heterologous DNA from the rapamycin or monensin PKSs can be exchanged for that native to the erythromycin PKS in order to generate novel polyketides with altered alkyl branching. Such an AT domain swap represented the first example of the production of a truly hybrid PKS (Oliynyk *et al.*, 1996). WO 98/01546 also describes in general terms the production of hybrid PKS assemblies comprising a loading module and at least one extension module. It specifically describes the construction of a hybrid PKS gene assembly by grafting the broad-specificity loading module for the avermectin-producing PKS onto the first protein of the erythromycin PKS (DEBS1) in place of the normal loading module (see also Marsden *et al.*, 1998). Additional examples comprising loading module swaps that are substrate specific have also been described (WO 00/00618; US 5876991; Kuhstoss *et al.*, 1996). WO 00/01827 describes methods for varying the β-keto processing capability of a PKS module through the ability to swap 'reductive loops', ie. the ability to rapidly and in a combinatorial manner, alter the number and type of ketoreductase, dehydratase and enoyl reductase domains within a module. In addition to changing the level of β-keto group processing, such changes may also lead to changes in stereochemistry of the α-alkyl and β-hydroxyl groups thus formed by the altered modules.

Although modular PKSs operate 'normally' in a co-linear and processive manner as described above, examples of a deviation from this mode of operation have been described and are discussed below.

The picromycin PKS gene cluster in *Streptomyces venezuelae* is responsible for the biosynthesis of both picromycin (a 14-membered, heptaketide macrolide) and methymycin (a 12-membered, hexaketide macrolide) (Xue *et al.*, 1998). The ability of a single PKS to produce two related macrolides, of different ring sizes, derives from the alternative expression of the final PKS gene *pikA4* (Xue & Sherman, 2000). When 'normal' expression occurs and full-length PikA4 is formed, a sixth extension unit is incorporated and the picromycin aglycone is produced; when alternative expression occurs and an N-terminally truncated form of PikA4 is produced, no sixth extension unit is incorporated and the growing polyketide chain is passed directly to the TE domain which leads to formation of the methymycin aglycone. Thus, a breakdown of co-linearity occurs and a 'ring contracted' product is formed. The biochemical basis for this phenomenon has been investigated and shown to be an ACP5 to ACP6 transfer, missing out covalent attachment to the intervening KS6 domain; such a breakdown of co-linearity has been called 'skipping' (Beck *et al.*, 2002).

Skipping has also been observed to occur when an extra extension module from the rapamycin PKS was interpolated into the erythromycin PKS in order to convert the natural heptketide-producing PKS into an octaketide-producing one (Rowe *et al.*, 2001). The expected octaketide, 16-membered macrolide was produced, but the major product was the normal heptaketide product 6-deoxyerythronolide. This 'skipping' of the interpolated module is believed to occur due to the interpolated module acting on some occasions as a 'shuttle', passing the growing chain from the preceding module to the following downstream module without performing a round of chain extension. It was subsequently shown that the ACP domain of the interpolated module is essential in passing the growing polyketide chain from the preceding ACP domain and passing it to the KS domain of the following module during skipping (Thomas *et al.*, 2002), a mechanism similar to that described for methymycin biosynthesis above. It is shown that skipping can occur without the active site nucleophile of the KS domain. A ring-contracted (skipped) nemadectin (an antiparasitic macrolide) has been reported from a mutant of a *Streptomyces* soil isolate that was modified by chemical mutation (Rudd *et al.*, 1990); the biosynthesis of the natural PKS product was abolished.

An alternative manner in which modular PKSs deviate from co-linear operation involves the iterative operation of modules. For example, module 4 of the erythromycin PKS appears to operate iteratively, at a low level, to produce a ring expanded 16-membered, octaketide macrolide related to 6-deoxyerythronolide B (Wilkinson *et al.*, 2000). The ability of the erythromycin PKS to perform this operation has been termed 'stuttering'. The 'stuttering' of the erythromycin PKS is considered an aberrant process, as the products of this stuttering are formed in low yield and the major product of the erythromycin PKS is the normal heptaketide 6-deoxyerythonolide B formed by co-linear operation. Products that appear to be formed by both stuttering and skipping have also been reported as minor components from the epothilone producer *Sorangium cellulosum* (Hardt *et al.*, 2001). The stigmatellin biosynthetic cluster of *Stigmatella aurantiaca* encodes for a PKS that comprises ten (one loading and nine extension) modules (Gaitatzis *et al.*, 2002); however, based on results from structural elucidation and the feeding of stable isotope labelled substrates, stigmatellin is formed from eleven modular derived units. Thus, it would appear that one of the stigmatellin PKS modules operates (twice) iteratively.

Since the priority filing of the present application, the sequence of the PKS responsible for biosynthesis of the macrolide lankacidin by *Streptomyces rochei* has been described (Mochizuki *et al.*, 2003). This PKS also appears to contain too few modules in comparison to the number of extension cycles required for lankacidin biosynthesis, although the mechanism by which this would occur is not clear.

Additional structural diversity can be generated through the modification of polyketides by enzymes other than the PKS, either during the process of chain assembly as seen during the biosynthesis of some ansamycins (Floss, 2001), or after the process of chain assembly following release from the PKS. Such non-PKS mediated reactions may include, but are not limited to the following: reduction, oxidation, hydroxylation, acylation, alkylation, amination, decarboxylation, dehydration, double bond isomerisation/migration, cyclisation, ring cleavage, conjugation, glycosylation, reductive elimination and any combination of these. When these reactions occur after chain assembly they are termed the post-PKS or tailoring steps. Such tailoring steps are generally, but not always, essential for endowing the polyketide natural product with biological activity.

In addition, the structural diversity of polyketides obtainable biosynthetically can be further enhanced through the use of defined heterologous post-PKS tailoring

enzymes as well as through the use of those which naturally modify the natural polyketide (Gaisser *et al.*, 2000). WO 01/79520 describes the heterologous modification of polyketide macrolide structures through glycosylation, epoxidation, hydroxylation, and methylation. The ability to generate analogues of the agricultural compound spinosyn through glycosylation with alternative deoxyhexose substituents has been reported (Gaisser *et al.*, 2002).

Borrelidin 1 (Figure 1) is an 18-membered macrolide produced by several bacterial strains including, but not limited to, *Streptomyces rochei* ATCC23956, *Streptomyces parvulus* Tü113 and *Streptomyces parvulus* Tü4055. Borrelidin is herein shown to be derived from a *trans*-cyclopentane-1,2-dicarboxylic acid starter acid, three malonyl-CoA and five methylmalonyl-CoA extender units (see figure 2). From the absolute stereochemistry of borrelidin, based on the crystal structure and recently confirmed through total synthesis, the actual starter acid is predicted to be *trans*-cyclopentane-(1*R*,2*R*)-dicarboxylic acid. Borrelidin isolated after the feeding of stable isotope labelled acetate and propionate substrates clearly indicated the expected incorporation of these building blocks; in addition, it has been demonstrated in the present application that feeding of *trans*-cyclopentane-1,2-dicarboxylic acid was sufficient to re-establish borrelidin biosynthesis in mutants where specific genes believed to be involved in the formation of the starter unit had been disrupted. Borrelidin contains a nitrile group attached to the C12 position, which is shown herein to arise through the action of tailoring enzymes acting upon a methylmalonyl-CoA derived methyl branch present at this position. The gross structure of borrelidin was first elucidated in 1967 (Keller-Scheirlein, 1967), and was subsequently refined by detailed NMR analysis (Kuo *et al.*, 1989). The absolute configuration of borrelidin was confirmed by X-ray crystallography (Anderson *et al.*, 1989). Its co-identity as the antibiotic treponemycin has been verified (Maehr & Evans, 1987).

A number of groups have reported the synthesis of fragments of the borrelidin structure, and since the priority filing of the present application, two independent total syntheses of borrelidin have been reported (Hanessian *et al.*, 2003; Duffey *et al.*, 2003).

Borrelidin was first discovered due to its antibacterial activity (Berger *et al.*, 1949), although this antibacterial activity extends only to a limited number of micrococci, and is not found against all common test bacteria. The mode of action in sensitive microorganisms involves selective inhibition of threonyl tRNA synthetase

(Paetz & Nass, 1973). Other activities against spirochetes of the genus *Treponema* (Singh *et al.*, 1985; US 4,759,928), against viruses (Dickinson *et al.*, 1965), uses for the control of animal pests and weeds (DE 3607287) and use as an agricultural fungicide (DE 19835669; US 6,193,964) have been reported. Additionally, since the priority filing of the present application, borrelidin has been reported to have antimalarial activity against drug resistant *Plasmodium falciparum* strains (Otoguro *et al.*, 2003). Between all of these reports only two reported any synthetically modified derivatives. The first of these describes the benzyl ester and its bis-*O*-(4-nitrobenzoyl) derivative (Berger *et al.*, 1949). The second of these describes the borrelidin methyl ester, the methyl ester bis O-acetyl derivative, and the methyl ester $\Delta_{14\text{-}15}$-dihydro-, $\Delta_{14\text{-}15,12\text{-}13}$-tetrahydro-, and $\Delta_{14\text{-}15,12\text{-}13}$-tetrahydro-C12-amino derivatives (Anderton & Rickards, 1965). No biological activity was reported for any of these compounds.

A recent disclosure of particular interest is the discovery that borrelidin displays anti-angiogenesis activity (Wakabayashi *et al.*, 1997). Angiogenesis is the process of the formation of new blood vessels. Angiogenesis occurs only locally and transiently in adults, being involved in, for example, repair following local trauma and the female reproductive cycle. It has been established as a key component in several pathogenic processes including cancer, rheumatoid arthritis and diabetic retinopathy. Its importance in enabling tumours to grow beyond a diameter of 1-2 cm was established by Folkman (Folkman, 1986), and is provoked by the tumour responding to hypoxia. In its downstream consequences angiogenesis is mostly a host-derived process, thus inhibition of angiogenesis offers significant potential in the treatment of cancers, avoiding the hurdles of other anticancer therapeutic modalities such as the diversity of cancer types and drug resistance (Matter, 2001). It is of additional interest that recent publications have described the functional involvement of tyrosinyl- and tryptophanyl tRNA synthetases in the regulation of angiogenesis (Wakasugi *et al.*, 2002; Otani *et al.*, 2002).

In the rat aorta matrix culture model of angiogenesis, borrelidin exhibits a potent angiogenesis-inhibiting effect and also causes disruption of formed capillary tubes in a dose dependent manner by inducing apoptosis of the capillary-forming cells (Wakabayashi *et al.*, 1997). Borrelidin inhibited capillary tube formation with an IC_{50} value of 0.4 ng/ml (0.8 nM). In the same study, borrelidin was shown to possess anti-proliferative activity towards human umbilical vein endothelial cells (HUVEC) in a cell growth assay; the IC_{50} value was measured at 6 ng/ml, which is 15-fold weaker than

the anti-angiogenesis activity measured in the same medium. This anti-proliferative activity of borrelidin was shown to be general towards various cell lines. In addition to these data the authors report that borrelidin inhibits tRNA synthetase and protein synthesis in the cultured rat cells; however the IC_{50} value for anti-angiogenesis activity (0.4 ng/ml) was 50-fold lower than that reported for inhibition of protein synthesis (20 ng/ml), indicating different activities of the compound.

Borrelidin also displays potent inhibition of angiogenesis *in vivo* using the mouse dorsal air sac model (Funahashi *et al.*, 1999), which examines VEGF-induced angiogenesis and is an excellent model for studying tumour-angiogenesis. Borrelidin was administered at a dose of 1.8 mg/kg by intraperitoneal injection and shown to significantly reduce the increment of vascular volume induced by WiDr cells, and to a higher degree than does TNP-470, which is a synthetic angiogenesis inhibitor in clinical trials. Detailed controls verified that these data are for angiogenesis inhibition and not inhibition of growth of the tumour cells. The authors also showed that borrelidin is effective for the inhibition of the formation of spontaneous lung metastases of B16-BL6 melanoma cells at the same dosage by inhibiting the angiogenic processes involved in their formation.

JP 9-227,549 and JP 8-173,167 confirm that borrelidin is effective against WiDr cell lines of human colon cancer, and also against PC-3 cell lines of human prostate cancer. JP 9-227,549 describes the production of borrelidin by *Streptomyces rochei* Mer-N7167 (Ferm P-14670) and its isolation from the resulting fermentation culture. In addition to borrelidin **1**, 12-desnitrile-12-carboxyl borrelidin **2** (presumably a biosynthetic intermediate or shunt metabolite), 10-desmethyl borrelidin **3** (presumably a biosynthetic analogue arising from the mis-incorporation of an alternative malonyl-CoA extender unit in module 4 of the borrelidin PKS), 11-epiborrelidin **4** and the C14,C15-cis borrelidin analogue **5** were described (see figure 1). Thus, JP 9-227,549 specifies borrelidin and borrelidin analogues wherein a nitrile or carboxyl group is attached the carbon skeleton at C12, and a hydrogen atom or lower alkyl group is attached to the carbon skeleton at C10.

WO 01/09113 discloses the preparation of borrelidin analogues that have undergone synthetic modification at the carboxylic acid moiety of the cyclopentane ring. The activity of these compounds was examined using endothelial cell proliferation and endothelial capillary formation assays in a similar manner to that described above. In general, modification of the carboxyl moiety improved the selectivity for inhibiting

capillary formation: the major reason for this improvement in selectivity is through a decrease in the cell proliferation inhibition activity whereas the capillary formation inhibitory activity was altered to a much lower degree. Specifically, the borrelidin-morpholinoethyl ester showed a 60-fold selectivity index, the borrelidin-amide showed a 37-fold selectivity index, the borrelidin-(2-pyridyl)-ethyl ester showed a 7.5-fold selectivity index and the borrelidin-morpholinoethyl amide showed a 6-fold selectivity index, for the capillary formation inhibitory activity versus cell proliferation with respect to borrelidin. The capillary formation inhibitory activity of these and other borrelidin derivatives was verified using a micro-vessel formation assay. In addition, the authors showed that borrelidin weakly inhibited the propagation of metastatic nodules, after removal of the primary tumour, when using a Lewis lung adenocarcinoma model. However, the borrelidin-(3-picolylamide) derivative was reported to inhibit very considerably the increase of micrometastases in rats after intraperitoneal and also with *per os* administration at subtoxic doses. Similarly, using the colon 38 spleen liver model, the metastasis-forming ability of mouse colon adenocarcinoma cells transplanted into mouse spleen was considerably decreased after treatment with a subtoxic dose of this borrelidin derivative. These data confirm the earlier reported ability of borrelidin and its derivatives to inhibit the formation of metastases.

Borrelidin has also been identified as an inhibitor of cyclin-dependant kinase Cdc28/Cln2 of *Saccharomyces cerivisiae* with an IC_{50} value of 12 μg/ml (24 μM) (Tsuchiya *et al.*, 2001). It was shown that borrelidin arrests both haploid and diploid cells in late G_1 phase (at a time point indistinguishable from α-mating pheromone), and at concentrations that do not affect gross protein biosynthesis. These data were taken to indicate that borrelidin has potential as a lead compound to develop anti-tumour agents.

Since the priority filing of the present application, two further reports have been published concerning the biological activity of borrelidin. The first of these indicates that the anti-angiogenic effects of borrelidin are mediated through distinct pathways (Kawamura *et al.*, 2003). High concentrations of threonine were found to attenuate the ability of borrelidin to inhibit both capillary tube formation in the rat aorta culture model and HUVEC cells proliferation; however, it did not affect the ability of borrelidin to collapse formed capillary tubes or to induce apoptosis in HUVEC. Borrelidin was also found to activate caspase-3 and caspase-8, and inhibitors of both of these suppressed borrelidin induced apoptosis in HUVEC. The second of these papers used the method

of global cellular mRNA profiling to provide insight into the effects of borrelidin on *Saccharomyces cerevisiae* (Eastwood and Schaus, 2003). This analysis showed the induction of amino acid biosynthetic enzymes in a time-dependent fashion upon treatment with borrelidin, and it was ascertained that the induction of this pathway involves the *GCN4* transcription factor.

In summary, the angiogenesis-inhibitory effect of borrelidin is directed towards the twin tumour-biological effects of proliferation and capillary formation. In addition, borrelidin, and derivatives thereof, have been shown to inhibit the propagation of metastases. Borrelidin also has indications for use in cell cycle modulation. Thus, borrelidin and related compounds are particularly attractive targets for investigation as therapeutic agents for the treatment of tumour tissues, either as single agents or for use as an adjunct to other therapies. In addition, they may be used for treating other diseases in which angiogenesis is implicated in the pathogenic process, including, but not restricted to, the following list: rheumatoid arthritis, psoriasis, atherosclerosis, diabetic retinopathy and various ophthalmic disorders.

SUMMARY OF THE INVENTION

The present invention provides the entire nucleic acid sequence of the biosynthetic gene cluster responsible for governing the synthesis of the polyketide macrolide borrelidin in *Streptomyces parvulus* Tü4055. Also provided is the use of all or part of the cloned DNA and the nucleic acid sequences thereof in the specific detection of other polyketide biosynthetic gene clusters, in the engineering of mutant strains of *Streptomyces parvulus* and other suitable host strains for the production of enhanced levels of borrelidin, or for the production of modified or novel polyketides, and of recombinant genes encoding PKS systems for the biosynthesis of modified or novel polyketides.

The present invention provides an isolated nucleic acid molecule comprising all or part of a borrelidin biosynthetic gene cluster.

The complete nucleotide sequence of the borrelidin biosynthetic gene cluster from *Streptomyces parvulus* Tü4055 is shown in SEQ ID No.1. Its organisation is presented in figure 3 and comprises genes and open reading frames designated hereinafter as: *borA1*, *borA2*, *borA3*, *borA4*, *borA5*, *borA6*, *borB*, *borC*, *borD*, *borE*, *borF*, *borG*, *borH*, *borI*, *borJ*, *borK*, *borL*, *borM*, *borN*, *borO*, *orfB1*, *orfB2*, *orfB3*, *orfB4*,

orfB5, *orfB6*, *orfB7*, *orfB8*, *orfB9*, *orfB10*, *orfB11*, *orfB12*, *orfB13*, *orfB14*, *orfB15*, *orfB16*, *orfB17*, *orfB18*, *orfB19*, *orfB20*, *orfB21* and *orfB22*.

The proposed functions of the cloned genes are described in Figures 4 (proposed biosynthesis of the starter unit), 5 (organisation of the borrelidin PKS and biosynthesis of pre-borrelidin) and 6 (introduction of the C12-nitrile moiety) and are described below.

The present invention thus provides an isolated nucleic acid molecule comprising:

(a) a nucleotide sequence as shown in SEQ ID No.1, or a portion or fragment thereof; or

(b) a nucleotide sequence which is the complement of SEQ ID No.1, or a portion or fragment thereof; or

(c) a nucleotide sequence which is degenerate with a coding sequence of SEQ ID No.1, or a portion or fragment thereof.

As used herein the term "fragment" with respect to nucleotide sequences refers to a stretch of nucleic acid residues that are at least 10, preferably at least 20, at least 30, at least 50, at least 75, at least 100, at least 150 or at least 200 nucleotides in length. A preferred portion or fragment of SEQ ID NO:1 is the sequence extending between nucleotide positions 7603 and 59966 of SEQ ID No.1.

The sequence may encode or be complementary to a sequence encoding a polypeptide of a polyketide biosynthetic gene cluster, or a portion thereof. By "a polypeptide of a polyketide biosynthetic gene cluster" is meant a polypeptide encoded by one or more open reading frames of a polyketide biosynthetic gene cluster, and particularly the borrelidin biosynthetic gene cluster.

A polyketide biosynthetic gene cluster is a segment of DNA comprising a plurality of genes encoding polypeptides having activity in the biosynthesis of a polyketide or macrolide moiety. This is not restricted to components of the polyketide synthase (PKS) which function *inter alia* in the synthesis of the polyketide backbone and reductive processing of side groups, but also encompasses polypeptides having ancillary functions in the synthesis of the polyketide. Thus polypeptides of the biosynthetic gene cluster may also act in macrolide ring or polyketide chain modification (e.g. catalysing a reaction in the formation of the C12 nitrile moiety of borrelidin), in the synthesis of a precursor or starter unit for a polyketide or macrolide moiety (e.g. catalysing a reaction in the synthesis of the *trans*-cyclopentane-1,2-

dicarboxylic acid starter unit for the borrelidin PKS, or responsible for the activation of such molecules as the coenzyme-A thioesters of the starter and extender units of the chain), regulatory activity (e.g. regulation of the expression of the genes or proteins involved in polyketide or macrolide synthesis), transporter activity (e.g. in transport of substrates for the polyketide or macrolide moiety into the cell, or of synthesis products such as the polyketide or macrolide molecule out of the cell), and in conferring resistance of the producing cell to the synthesised products (e.g. through specific binding to the synthesised molecule, or as a replacement for other endogenous proteins to which the synthesised molecule may bind within or outside of the cell).

The gene cluster also includes non-coding regions, such as promoters and other transcriptional regulatory sequences which are operably linked to the coding regions of the gene cluster. The skilled person is well able to identify such elements based upon the information provided herein, and these are within the scope of the present invention.

Genes and open reading frames encoded within SEQ ID No.1 represent preferred parts or fragments of SEQ ID No.1. Thus an isolated nucleic acid molecule may comprise a sequence that encodes a polypeptide from a borrelidin biosynthetic gene cluster, wherein said polypeptide has an amino acid sequence selected from the group consisting of SEQ ID Nos.2 to 43 and 113.

In preferred embodiments, the nucleic acid sequence comprises an open reading frame selected from the group of open reading frames of SEQ ID NO: 1 consisting of *borA1*, *borA2*, *borA3*, *borA4*, *borA5*, *borA6*, *borB*, *borC*, *borD*, *borE*, *borF*, *borG*, *borH*, *borI*, *borJ*, *borK*, *borL*, *borM*, *borN*, *borO*, *orfB1*, *orfB2*, *orfB3*, *orfB4*, *orfB5*, *orfB6*, *orfB7*, *orfB8*, *orfB9*, *orfB10*, *orfB11*, *orfB12*, *orfB13a*, *orfB13b*, *orfB14*, *orfB15*, *orfB16*, *orfB17*, *orfB18*, *orfB19*, *orfB20*, *orfB21* and *orfB22*, said open reading frames being described by, respectively, bases 16184*-18814, 18875-23590, 23686-34188, 34185*-39047, 39122*-45514, 45514-50742, 7603-8397c, 8397-9194c, 9244-9996c, 9993-11165c, 11162-11980c, 11992-13611c, 13608-15659*c, 50739*-52019, 52113-53477, 53486-54466, 54506-56176, 56181*-57098, 57112-57858, 57939-59966, 2-313 (incomplete), 501*-3107, 3172-3810c, 3935-4924c, 5123-5953, 5961-6518*c, 6564*-7538, 60153-60533*c, 60620-61003, 61188*-61436, 61526-61738, 61767-62285c, 62750-63067c, 62586-62858c, 63155-65071c, 65374-65871, 65942-68305*c, 68290-68910*c, 69681-70436, 70445-71848, 71851-72957, 73037-73942 and 73995-74534c of SEQ ID No.1.

In the above list, 'c' indicates that the gene is encoded by the complementary strand to that shown in SEQ ID NO: 1. Each open reading frame above represents the longest probable open reading frame present. It is sometimes the case that more than one potential start codon can be identified. One skilled in the art will recognise this and be able to identify alternative possible start codons. Those genes which have more than one possible start codon are indicated with a '*' symbol. Throughout we have indicated what we believe to be the start codon, however, a person of skill in the art will appreciate that it may be possible to generate active protein using an alternative start codon, proteins generated using these alternative start codons are also considered within the scope of the present invention.

It should be noted that a number of these open reading frames begin with a codon (GTG, CTG or TTG) other than the more normal ATG initiation codon. It is well known that in some bacterial systems such codons, which normally denote valine (GTG) or leucine (CTG, TTG), may be read as initiation codons encoding methionine at the N terminus of the polypeptide chain. In the amino acid sequences (SEQ ID Nos: 2 to 43 and 113) provided herein, such codons are therefore translated as methionine.

Also provided are nucleic acid molecules comprising portions of the open reading frames identified herein. For example, such a nucleic acid sequence may comprise one or more isolated domains derived from the open reading frames identified herein. The polypeptides encoded by these isolated portions of the open reading frames may have independent activity, e.g. catalytic activity. In particular, the polypeptides which make up the borrelidin PKS have modular structures in which individual domains have particular catalytic activities as set out above. Thus any of these domains may be expressed alone or in combination, with other polypeptides from the borrelidin PKS described herein or domains thereof, or with polypeptides from the PKS of other polyketides. In particular, any of these domains may be substituted for the equivalent domains either within the borrelidin PKS or in other polyketide synthases and additionally equivalent domains from other PKSs may be substituted for domains within the borrelidin PKS. In this context an equivalent domain includes domains which have the same type of function but differ in for example, their specificity, an example of substitutions contemplated by the present invention include: the substitution of a malonyl-CoA specific AT domain for a methylmalonyl-CoA specific AT domain, or the substitute of a reductive loop containing a KR domain only for one containing KR, DH

and ER. In preferred embodiments the expressed domains represent at least one PKS module as described below.

The term 'PKS domain' as used herein refers to a polypeptide sequence, capable of folding independently of the remainder of the PKS, and having a single distinct enzymatic activity or other function in polyketide or macrolide synthesis including, but not restricted to β-ketoacyl-acyl carrier protein synthase (KS), acyl carrier protein (ACP), acyl transferase (AT), β-ketoreductase (KR), dehydratase (DH), enoyl reductase (ER) or terminal thioesterase (TE).

Accordingly, the invention further provides:

(a) an isolated nucleic acid molecule comprising a sequence that encodes a PKS domain selected from AT0 and ACP0, said domains being described by, respectively, amino acids 322-664 and 694-763 of SEQ ID No.2. In a preferred embodiment, the PKS domain comprises a sequence selected from the group consisting of bases 17147-18175 and 18263-18472 of SEQ ID No.1;

(b) an isolated nucleic acid molecule comprising a sequence that encodes a PKS domain selected from KS1, AT1, KR1 and ACP1, said domains being described by, respectively, amino acids 34-459, 557-885, 1136-1379 and 1419-1486 of SEQ ID No.3. In a preferred embodiment, the PKS domain comprises a sequence selected from the group consisting of bases 18974-20251, 20543-21529, 22280-23011 and 23129-23332 of SEQ ID No.1;

(c) an isolated nucleic acid molecule comprising a sequence that encodes a PKS domain selected from KS2, AT2, DH2, KR2, ACP2, KS3, AT3, DH3, KR3 and ACP3, said domains being described by, respectively, amino acids 34-459, 559-887, 903-1050, 1354-1597, 1628-1694, 1724-2149, 2245-2576, 2593-2734, 3060-3307 and 3340-3406 of SEQ ID No.4. In a preferred embodiment, the PKS domain comprises a sequence selected from the group consisting of bases 23785-25062, 25360-26346, 26392-26835, 27745-28476, 28567-28767, 28855-30132, 30418-31413, 31462-31887, 32863-33606 and 33703-33903 of SEQ ID No.1;

(d) an isolated nucleic acid molecule comprising a sequence that encodes a PKS domain selected from KS4, AT4, KR4 and ACP4, said domains being described by, respectively, amino acids 34-459, 555-886, 1179-1423 and 1459-1525 of SEQ ID No.5. In a preferred embodiment, the PKS domain comprises a sequence selected from the group consisting of bases 34284-35561, 35847-36842, 37719-38453 and 38559-38759 of SEQ ID No.1;

(e) an isolated nucleic acid molecule comprising a sequence that encodes a PKS domain selected from KS5, AT5, DH5, ER5, KR5 and ACP5, said domains being described by, respectively, amino acids 34-457, 553-888, 905-1046, 1401-1690, 1696-1942 and 1975-2041 of SEQ ID No.6. In a preferred embodiment, the PKS domain comprises a sequence selected from the group consisting of bases 39221-40492, 40778-41785, 41834-42259, 43322-44191, 44207-44947 and 45044-45244 of SEQ ID No.1;

(f) an isolated nucleic acid molecule comprising a sequence that encodes a PKS domain selected from KS6, AT6, KR6, ACP6 and TE, said domains being described by, respectively, amino acids 37-457, 555-883, 1101-1335, 1371-1437 and 1461-1708 of SEQ ID No.7. In a preferred embodiment, the PKS domain comprises a sequence selected from the group consisting of bases 45622-46884, 47176-48162, 48814-49518, 49624-49824 and 49894-50637 of SEQ ID No.1.

In another of its aspects the invention provides an isolated nucleic acid molecule comprising a sequence that encodes a PKS module, said module being selected from the group consisting of amino acids 322-763 of SEQ ID No.2, 34-1486 of SEQ ID No.3, 34-1694 of SEQ ID No.4, 1724-3406 of SEQ ID No.4, 34-1525 of SEQ ID No.5, 34-2041 of SEQ ID No.6 and 37-1437 or 1708 of SEQ ID No.7. In a preferred embodiment, the module comprises a sequence selected from the group consisting of bases 17147-18472, 18974-23332, 23785-28767, 28855-33903, 34284-38759, 39221-45244, 45622-49824 or 50637 of SEQ ID No.1.

The term 'module' as used herein refers to a single polypeptide comprising a plurality of PKS domains each having a single distinct enzymatic activity in polyketide or macrolide synthesis including, but not restricted to β-ketoacyl-acyl carrier protein synthase (KS), acyltransferase (AT), acyl carrier protein (ACP), β-ketoreductase (KR), dehydratase (DH), or enoyl reductase (ER) or terminal thioesterase (TE). An extension module typically comprises a KS, AT and ACP domain (although some modular PKSs may encode their AT domains as independent proteins). An extension module may further comprise one or more domains capable of reducing a beta-keto group to a hydroxyl, enoyl or methylene group (said group of domains are referred to herein as a "reductive loop"). Thus a module comprising a reductive loop typically contains a KR domain, KR and DH domains, or KR, DH and ER domains.

A PKS may further comprise a TE domain to perform chain termination and/or cyclisation of the final product, or alternatively it may contain another functionality

known to perform a similar function such as that for the addition of an amino acid residue and macrolactam formation as observed for rapamycin (Schwecke et al., 1995), for macrolactam formation as for rifamycin (August et al., 1998), and for amino acid incorporation followed by reductive elimination as for myxalamid biosynthesis (Silakowski et al., 2001)..

Also provided is a nucleic acid molecule encoding a polyketide synthase comprising a sequence encoding one or more of the domains or modules described above.

The sequences provided herein provide means with which to manipulate and/or to enhance polyketide synthesis. Thus there is provided a method of modifying a parent polyketide synthase, comprising expressing a domain from a borrelidin polyketide synthase or a derivative thereof as described herein in a host cell expressing said parent polyketide synthase, such that the domain is incorporated into said parent polyketide synthase. There is further provided a method of modifying a parent polyketide synthase, comprising introducing into a host cell a nucleic acid encoding a domain from a borrelidin polyketide synthase, or a derivative thereof, wherein the host cell contains nucleic acid encoding said parent polyketide synthase, such that, when expressed, the domain is incorporated into said parent polyketide synthase.The borrelidin PKS domain may be inserted in addition to the native domains of the parent PKS, or may replace a native parent domain. Typically the parent PKS will be a Type I PKS.

The present invention further provides methods of modifying a parent borrelidin PKS. A donor domain (e.g. from a Type I PKS) may be expressed in a host cell expressing said parent borrelidin PKS. There is further provided a method of modifying a parent borrelidin polyketide synthase comprising introducing into a host cell a nucleic acid encoding a domain from a donor polyketide synthase, wherein the host cell contains nucleic acid encoding said parent borrelidin polyketide synthase, such that, when expressed, the domain is incorporated into said parent borrelidin polyketide synthase.

Additionally or alternatively, a domain of the parent PKS may be deleted or otherwise inactivated; e.g. a parent domain may simply be deleted, or be replaced by a domain from a donor PKS, or a domain from a donor PKS may be added to the parent. Where a domain is added or replaced, the donor domain may be derived from the parent synthase, or from a different synthase.

These methods may be used to enhance the biosynthesis of borrelidin, to produce new borrelidin derivatives or analogues, or other novel polyketide or macrolide structures. The number and nature of modules in the system may be altered to change the number and type of extender units recruited, and to change the various synthase, reductase and dehydratase activities that determine the structure of the polyketide chain. Such changes can be made by altering the order of the modules that comprise the PKS, by the duplication or removal of modules that comprise the PKS, by the introduction of modules from heterologous sources, or by some combination of these various approaches.

Thus domains or modules of the borrelidin PKS may be deleted, duplicated, or swapped with other domains or modules from the borrelidin PKS, or from PKS systems responsible for synthesis of other polyketides (heterologous PKS systems, particularly Type I PKS systems), which may be from different bacterial strains or species. Alternatively domains or modules from the borrelidin PKS may be introduced into heterologous PKS systems in order to produce novel polyketide or macrolides. Combinatorial modules may also be swapped between the borrelidin polyketide synthase and other polyketide synthases, these combinatorial modules extend between corresponding domains of two natural-type modules, e.g. from the AT of one module to the AT of the next.

For example, a particular extender module may be swapped for one having specificity for a different extender unit (as described e.g. in WO98/01571 and WO98/01546), or mutated to display specificity or selectivity for a different extender unit e.g. as described below. Additionally or alternatively, introduction, deletion, swapping or mutation of domains or modules, such as the KR, DH and ER domains responsible for the processing of a given β-keto moiety, may be used to alter the level of reductive processing of an extender unit during polyketide synthesis. Such changes may also lead to changes in stereochemistry of the alpha-alkyl and beta-hydroxyl groups thus formed by altered modules. In a preferred embodiment the BorA5 module may be introduced into a parent PKS to provide iterative addition of extender units to a polyketide backbone, e.g. expanding the ring size of a macrolide polyketide relative to that naturally produced by the parent PKS.

The borrelidin loading module is the first PKS loading module to be identified having specificity for an alicyclic di-carboxylic acid starter unit. Thus this module or a derivative thereof may be used to introduce alicyclic starter units into heterologous

polyketide synthases. This need not be restricted to use of *trans*-cyclopentane-1,2,-dicarboxylic acid normally used as the borrelidin starter unit. The borrelidin loading module is herein shown also to be capable of directing incorporation of other starter units including *trans*-cyclobutane-1,2-dicarboxylic acid, 2,3-methylsuccinic acid and 2-methylsuccinic acid. The borrelidin starter unit may also be modified in a borrelidin producing cell, or replaced by a heterologous loading module, to introduce alternative starter units into the borrelidin synthetic pathway.

The position of the loading module of the PKS may be chosen (e.g. by fusing it to a particular location within the PKS) in order to control the ring size of the resultant polyketide/macrolide molecules.

The AT domains that determine the carboxylic acid-CoA thioester extender units may be deleted, modified or replaced. The ACP domains may also be deleted, modified or replaced. In addition domains that are not normally present in the borrelidin PKS but which are found in other modular PKS and/or mixed PKS/NRPS systems may be inserted. Examples include, but are not limited to: O-methyl transferase domains, C-methyl transferase domains, epimerisation domains, monooxygenase domains and dehydrogenase domains, aminotransferase domains and non-ribosomal peptide synthetase domains.

Further, the thioesterase domain of the borrelidin PKS may be altered or repositioned (e.g. fused to a chosen location within the PKS) in order to change its specificity and/or in order to release polyketide/macrolide molecules with a chosen ring size. Alternatively, heterologous thioesterase domains may be inserted into the borrelidin PKS to produce molecules with altered ring size relative to the molecule normally produced by the parent PKS, or to produce a free acid.

In yet another alternative, the amino acid incorporating and macrolactam-forming domains from mixed NRPS/PKS systems such as that for rapamycin, or for related systems such as for rifamycin biosynthesis and myxalamid biosynthesis, or modules from NRPS systems (such as those for bleomycin biosynthesis) may be inserted into the PKS to produce novel polyketide related molecules of mixed origin.

The open reading frames encoding the PKS described herein may also comprise portions encoding non-enzymatically active portions which nevertheless have a functional role as scaffold regions which space and stabilise the enzymatically active domains and/or modules of the PKS at appropriate distances and orientations, and which may have recognition and docking functions that order the domains and modules

of the PKS in the correct spatial arrangement. Thus the nucleic acid sequences of the present invention comprise sequences encoding such scaffold regions, either alone or in combination with sequences encoding domains or modules as described above.

It will be appreciated that the various manipulations of PKS coding sequences described above may give rise to hybrid PKS genes or systems. Thus the present invention also provides nucleic acids encoding such hydrid PKS systems. The invention therefore provides a nucleic acid construct comprising at least one first nucleic acid portion encoding at least one domain of a borrelidin PKS and a second nucleic acid portion or portions encoding at least one type I PKS domain which is heterologous to said borrelidin PKS. In preferred embodiments the construct comprises a hybrid polyketide synthase gene, said gene encoding at least one domain of a borrelidin PKS and at least one type I PKS domain which is heterologous to said borrelidin PKS. Further preferred embodiments are as described above.

In a further aspect, the present invention provides an isolated nucleic acid molecule comprising a sequence encoding a polypeptide which catalyses a step in the synthesis of a starter unit or substrate for polyketide synthesis, preferably in the synthesis of the *trans*-cyclopentane-1,2,-dicarboxylic acid moiety used as a starter unit by the borrelidin PKS. The polypeptide may have activity as a dehydrogenase, 3-oxoacyl-ACP-reductase, cyclase, F420 dependent dehydrogenase, or 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase. Preferably the polypeptide comprises the sequence encoded by one of the group of genes consisting of *borC, borD, borE, borF, borG, borH,* borK, borL, *borM* and *borN*, as shown in SEQ ID NO: 8, 9, 10, 12, 13, 14, 17, 18, 19 or 20.

These genes may be rendered deleted, disrupted, or otherwise inactivated in a borrelidin-producing cell in order to abolish borrelidin production. Cell lines resulting from such changes may be chemically complemented by the addition of exogenous carboxylic acids which may be incorporated in place of the natural starter unit. Thus, new borrelidin related molecules may be synthesised, which are initiated from the exogenously fed carboxylic acid. Such an approach is termed mutasynthesis. The genes responsible for *trans*-cyclopentane-1,2,-dicarboxylic acid synthesis may be introduced into a heterologous polyketide producer cell to allow that cell to synthesise the alicyclic dicarboxylic acid as a starter unit for its own PKS.

Thus the present invention further provides a method for the production of borrelidin and borrelidin analogues at improved titres, said method comprising

disrupting *borG* in the host strain, fermenting the resulting cell line and feeding an exogenous carboxylic acid. In various preferred embodiments the exogenous carboxylic acid is *trans*-cyclopentane-1,2-dicarboxylic acid or the exogenous carboxylic acid is selected from the group consisting of *trans*-cyclobutane-1,2-dicarboxylic acid, 2,3-dimethyl succinic acid and 2-methylsuccinic acid and /or the method additional comprises deleting, modifing or replacing one or more borrelidin biosynthetic genes, or borrelidin polyketide synthase domains or modules. A person of skill in the art is aware that polyketide synthases may also be expressed in heterologous hosts, therefore the present invention also contemplates a method for the production of higher titres of borrelidin and borrelidin analogues in a heterologous host, said method comprising transforming a host cell with the entire borrelidin gene cluster with the exception of *borG* or disrupting the *borG* gene in situ once the gene cluster has been transferred.

Alternatively, genes responsible for the synthesis of the starter unit may be over-expressed in order to improve the fermentation titres of borrelidin or borrelidin related molecules. Thus the present invention further provides a method for increasing the titre of borrelidin and borrelidin derivatives or borrelidin related molecules and their derivatives, said method comprising upregulating a borrelidin biosynthetic gene involved in production of the starter unit, said gene selected from the group consisting of *borC, borD, borE, borF,* borH, *borK, borL borM* and borN, in a preferred embodiment the upregulated gene is *borE* or *borL*.

In another approach the genes responsible for the synthesis of the starter unit may be modified, or replaced by other synthetic genes directing the production of altered carboxylic acids, leading to the production of borrelidin related molecules. These techniques may be complemented by the modification of the loading module of the PKS as described above.

In a further aspect, the present invention provides an isolated nucleic acid molecule comprising a sequence encoding a polypeptide which catalyses a step in the modification of a side chain of a polyketide moiety, for example in the conversion of a methyl group to a nitrile moiety, e.g. at C12 of pre-borrelidin (**14**). The polypeptide may have activity as a cytochrome P450 oxidase, amino transferase, or NAD/quinone oxidoreductase. Preferably the polypeptide comprises the sequence encoded by one of the group of genes consisting of *borI, borJ,* and *borK* as shown in SEQ ID NO: 15, 16 or 17.

Various of these genes may be deleted/inactivated such that borrelidin-related molecules, or shunt metabolites thereof, accumulate which represent intermediate stages of the process that introduces the nitrile moiety. The addition of heterologous genes to such systems may allow alternative elaboration of any accumulated biosynthetic intermediates or shunt metabolites thereof. Alternatively, the genes may be mutated in order to alter their substrate specificity such that they function on alternative positions of pre-borrelidin molecules in order to provide borrelidin-related molecules. In addition, the genes responsible for formation of the nitrile group may be over-expressed in order to improve the fermentation titres of borrelidin or borrelidin-related molecules.

Alternatively, one, some or all of these genes may be introduced into cells capable of producing other polyketides to provide for desired side chain processing of that polyketide, e.g. the introduction of a nitrile moiety. This opens up the possibility of specific biosynthetic introduction of nitrile moieties into polyketides, particularly at side chains derived from methylmalonyl-CoA or ethylmalonyl-CoA extender units. Purified enzymes (see below) may also be used to effect the conversion of polyketide side chains to nitrile moieties *in vitro*.

In a further aspect, the present invention provides an isolated nucleic acid molecule comprising a sequence encoding a polypeptide involved conferring resistance to borrelidin. The polypeptide may have homology to a threonyl tRNA synthase, and preferably has threonyl tRNA synthase activity. Preferably the polypeptide comprises the sequence encoded by the *borO* gene as shown in SEQ ID NO: 21. A resistance gene such as *borO*, carried on a suitable vector (see below) may be used as a selective marker. Thus cells transformed with such a vector may be positively selected by culture in the presence of a concentration of borrelidin which inhibits the growth of, or kills, cells lacking such a gene.

In a further aspect, the present invention provides an isolated nucleic acid molecule comprising a sequence encoding a polypeptide involved in regulation of expression of one or more genes of the borrelidin gene cluster. In a preferred embodiment the polypeptide comprises the sequence encoded by the *borL* gene as shown in SEQ ID NO: 18, or as encoded by *orfB8* or *orfB12* as shown in SEQ ID NO: 29 or 33. Regulator genes may be engineered to increase the titre of borrelidin and borrelidin derivatives, or borrelidin related molecules and their derivatives produced by fermentation of the resulting cell lines. For example, repressors may be

deleted/inactivated, and/or activators may be up-regulated or overexpressed, e.g. by increasing gene copy number or placing the coding sequence under the control of a strong constitutively active or inducible promoter. The *borL* gene or a portion thereof may also find use as a hybridisation probe to identify similar regulator genes located in or outside other biosynthetic gene clusters.

In a further aspect, the present invention provides an isolated nucleic acid molecule comprising a sequence encoding a polypeptide having type II thioesterase activity. In a preferred embodiment the polypeptide comprises the sequence encoded by the *borB* gene as shown in SEQ ID NO: 8. This nucleic acid may be introduced into a host cell to modulate the titre of a polyketide synthesised by that cell. In particular, the titre may be increased by 'editing' of the products of unwanted side reactions (e.g. removal of acyl groups formed by inappropriate decarboxylation of extender units attached to KS domains). However in various aspects it may be desirable to remove such an activity from a producer cell, for example to increase the variety of polyketide products produced by that cell, or to facilitate production of an analogue of a naturally produced polyketide which would normally be blocked by such an editing activity.

The nucleotide sequences of the invention may be portions of the sequence shown in SEQ ID NO: 1, or the complement thereof, or mutants, variants, derivatives or alleles of these sequences. The sequences may differ from that shown by a change which is one or more of addition, insertion, deletion and substitution of one or more nucleotides of the sequence shown. Changes to a coding nucleotide sequence may result in an amino acid change at the protein level, or not, as determined by the redundancy of the genetic code. Thus, nucleic acid according to the present invention may include a sequence different from the sequence shown in SEQ ID NO: 1 yet encode a polypeptide with the same amino acid sequence. Preferably mutants, variants, derivatives or alleles of the sequences provided encode polypeptides having the same enzymatic activity as those described herein.

Where the sequence is a coding sequence, the encoded polypeptide may comprise an amino acid sequence which differs by one or more amino acid residues from the amino acid sequences shown in SEQ ID Nos: 2 to 43 and 113. Nucleic acid encoding a polypeptide which is an amino acid sequence mutant, variant, derivative or allele of any of the sequences shown is further provided by the present invention. Such polypeptides are discussed below. Nucleic acid encoding such a polypeptide may show greater than about 60% identity with the coding sequence of SEQ ID NO: 1,

greater than about 70% identity, greater than about 80% identity, or greater than about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity therewith. Percentage identity may be calculated using one of the programs such as BLAST or BestFit from within the Genetics Computer Group (GCG) Version 10 software package available from the University of Wisconsin, using default parameters.

In preferred embodiments, whether coding or non-coding, the nucleotide sequences of the invention are capable of hybridising specifically with at least a portion of the sequence of SEQ ID NO: 1 or the complement thereof.

For example, hybridizations may be performed, according to the method of Sambrook *et al.* (Sambrook *et al.*, 1989), using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 0.5-1.0% SDS, 100 μg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 1X SSC and 1% SDS; (4) 2 hours at 42-65°C in 1X SSC and 1% SDS, changing the solution every 30 minutes.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology is (Sambrook *et al.*, 1989):

T_m = 81.5°C + 16.6Log [Na+] + 0.41(% G+C) - 0.63 (% formamide) - 600/#bp in duplex

As an illustration of the above formula, using [Na+] = [0.368] and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C. Such hybridisation would be considered substantially specific to the nucleic acid sequence of the present invention.

The nucleic acids of the present invention preferably comprise at least 15 contiguous nucleotides of SEQ ID NO: 1. They may comprise 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 150, 200, 300, 500 or more contiguous nucleotides of SEQ ID NO: 1.

The nucleic acids may be used e.g. as primers or probes for the identification of novel genes or other genetic elements, such as transcriptional regulatory sequences, from polyketide or macrolide biosynthetic gene clusters, e.g. sequences encoding

enzymes of the PKS, or domains or modules thereof, enzymes involved in the biosynthesis of a starter unit, enzymes modifying side chains of polyketide moieties, transporters, resistance genes and regulatory molecules as described.

Thus the present invention provides a method of identifying a novel polyketide biosynthetic gene cluster, or a portion thereof, comprising hybridising a sample of target nucleic acid with a nucleic acid of the present invention capable of hybridising specifically to a nucleic acid having the sequence of SEQ ID NO: 1 or a portion thereof. The target nucleic acid may be any suitable nucleic acid, and is preferably bacterial genomic DNA.

Typically, the method further comprises the step of detecting hybridisation between the sample of nucleic acid and the nucleic acid of the invention. Hybridisation may be measured using any of a variety of techniques at the disposal of those skilled in the art. For instance, probes may be radioactively, fluorescently or enzymatically labelled. Other methods not employing labelling of probe include amplification using PCR, RNAase cleavage and allele specific oligonucleotide probing.

A method may include hybridization of one or more (e.g. two) probes or primers to target nucleic acid. Where the nucleic acid is double-stranded DNA, hybridization will generally be preceded by denaturation to produce single-stranded DNA. The hybridization may be as part of a PCR procedure, or as part of a probing procedure not involving PCR. An example procedure would be a combination of PCR and low stringency hybridization. A screening procedure, chosen from the many available to those skilled in the art, is used to identify successful hybridization events and isolated hybridized nucleic acid.

Those skilled in the art are well able to employ suitable conditions of the desired stringency for selective hybridisation, taking into account factors such as oligonucleotide length and base composition, temperature and so on, as described above.

An isolated nucleic acid molecule of the invention may be an isolated naturally occurring nucleic acid molecule (i.e. isolated or separated from the components with which it is normally found in nature) such as free or substantially free of nucleic acid flanking the gene in the bacterial genome, except possibly one or more regulatory sequence(s) for expression. Nucleic acid may be wholly or partially synthetic and may include genomic DNA, cDNA or RNA. Where nucleic acid according to the invention

includes RNA, reference to the sequence shown should be construed as reference to the RNA equivalent, with U substituted for T.

The present invention further provides a vector comprising a nucleic acid according to the present invention. The vector is preferably an expression vector comprising a nucleic acid encoding a polypeptide of a polyketide biosynthetic gene cluster (preferably a borrelidin biosynthetic gene cluster), or a portion thereof, as described. Suitable vectors comprising nucleic acid for introduction into bacteria or eukaryotic host cells can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, enhancer sequences, marker genes and other sequences as appropriate. Vectors may be plasmids, viral eg "phage", or "phagemid", as appropriate. For further details see, for example, Sambrook *et al.*, 1989. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in Short Protocols in Molecular Biology, Second Edition, Ausubel *et al.* Eds, John Wiley & Sons 1992. The disclosures of Sambrook *et al.* and Ausubel *et al.* are incorporated herein by reference.

In another of its aspects the present invention provides an isolated polypeptide encoded by a nucleic acid molecule of the invention as described herein. More particularly, there is provided an isolated polypeptide comprising an amino acid sequence as shown in any one or more of SEQ ID Nos.2 to 43 and 113 or a portion thereof. As set out above, these amino acid sequences represent translations of the longest possible open reading frames present in the sequence of SEQ ID NO: 1 and the complement thereof. The first amino acid is always shown as Met, regardless of whether the initiation codon is ATG, GTG, CTG or TTG.

As used herein the term "polypeptide(s)" includes peptides, polypeptides and proteins, these terms are used interchangeably unless otherwise specified.

A polypeptide which is an amino acid sequence variant, allele, derivative or mutant of any one of the amino acid sequences shown may exhibit at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity with the polypeptide of any one of the SEQ ID Nos.2 to 43 and 113, or with a portion thereof. Particular amino acid sequence variants may differ from those shown by insertion, addition, substitution or deletion of 1 amino acid, 2, 3, 4, 5-10, 10-20 20-30, 30-50, 50-100, 100-150, or more than 150 amino acids. Percentage identity may

be calculated using one of the programs such as FASTA or BestFit from within the Genetics Computer Group (GCG) Version 10 software package available from the University of Wisconsin, using default parameters.

The present invention also includes active portions, fragments, and derivatives of the polypeptides of the invention.

An "active portion" means a peptide which is less than the full length polypeptide, but which retains at least some of its essential biological activity. For example, isolated domains or modules of the PKS as described above may be regarded as active portions of the PKS

A "fragment" means a stretch of amino acid residues of at least five, at least six, or at least seven contiguous amino acids, often at least eight or at least nine contiguous amino acids, typically at least 10, at least 13 contiguous amino acids and, most preferably, at least 20, at last 25, at least 30, at least 50, at least 75, at least 100 or more contiguous amino acids. Fragments of the sequence may comprise antigenic determinants or epitopes useful for raising antibodies to a portion of the relevant polypeptide. Thus the polypeptide need not comprise a complete sequence provided in any one of SEQ ID Nos 2 to 43 and 113, but may comprise a portion thereof having the desired activity, e.g. an isolated domain or module, such as those of the PKS described above. It should be noted that the terms part, portion and fragment are used interchangeably in this specification; no particular significance should be ascribed to the specific use of one of these terms in any particular context.

A "derivative" of a polypeptide of the invention or a fragment thereof means a polypeptide modified by varying the amino acid sequence of the protein, e.g. by manipulation of the nucleic acid encoding the protein or by altering the protein itself. Such derivatives of the natural amino acid sequence may involve insertion, addition, deletion or substitution of one, two, three, five or more amino acids, without fundamentally altering the essential activity of the wild type polypeptide.

Polypeptides of the invention are provided in isolated form, e.g. isolated from one or more components with which they are normally found associated in nature. They may be isolated from a host in which they are naturally expressed, or may be synthetic or recombinant.

The present invention also encompasses a method of making a polypeptide (as disclosed), the method including expression from nucleic acid encoding the polypeptide (generally nucleic acid according to the invention). This may conveniently be achieved

by growing a host cell in culture, containing an expression vector as described above, under appropriate conditions which cause or allow expression of the polypeptide. Polypeptides may also be expressed in *in vitro* systems, such as reticulocyte lysate systems.

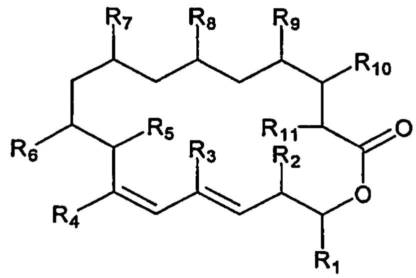
The method may include the step of introducing the nucleic acid into a host cell. The introduction, which may (particularly for *in vitro* introduction) be generally referred to without limitation as "transformation", may employ any available technique. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, conjugation, electroporation and transfection using bacteriophage. As an alternative, direct injection of the nucleic acid could be employed. Marker genes such as antibiotic resistance or sensitivity genes may be used in identifying clones containing nucleic acid of interest, as is well known in the art.

Preferred host cells include Actinomycetes, preferably Streptomycetes, and in particular those selected from the group consisting of *Saccharopolyspora erythraea, Streptomyces coelicolor, Streptomyces avermitilis, Streptomyces griseofuscus, Streptomyces cinnamonensis, Micromonospora griseorubida, Streptomyces hygroscopicus, Streptomyces fradiae, Streptomyces longisporoflavus, Streptomyces lasaliensis, Streptomyces tsukubaensis, Streptomyces griseus, Streptomyces venezuelae, Streptomyces antibioticus, Streptomyces lividans, Streptomyces rimosus and Streptomyces albus. Streptomyces rochei* ATCC23956, *Streptomyces parvulus* Tü113 and *Streptomyces parvulus* Tü4055, more preferably selected from the group consisting of *Streptomyces rochei* ATCC23956, *Streptomyces parvulus* Tü113 and *Streptomyces parvulus* Tü4055.

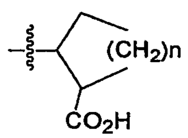
A polypeptide, peptide fragment, allele, mutant or variant according to the present invention may be used as an immunogen or otherwise in obtaining specific antibodies, which may be useful in purification and other manipulation of polypeptides and peptides, screening or other applications.

In another of its aspects the invention provides for the molecules that may be derived from the objects of the invention and for modified compounds formed therefrom and for methods for their production. The molecules derived from the objects of

the invention are shown by formula 1 and extends to pharmaceutically acceptable salts thereof, wherein:

Formula 1 

R_1 is a cycloalkyl group of varying size (n = 1- 2) and substituted as shown below;



wherein R_1 can also optionally be substituted with one or more halo atoms, or one or more C_1 to C_3 alkyl groups; R_2, R_3, R_6, R_7, R_8, R_9, or R_{11} are each independently H, OCH_3, CH_3 or CH_2CH_3; R_4 is CN, CO_2H, CHO, CH_3, $CONH_2$, CHNH, R_5, R_{10} are OH; or analogues differing from the corresponding "natural" compound in the oxidation state of one or more of the ketide units as shown in Figure 2 (i.e. selection of alternatives from the group: -CO-, -CH(OH)-, =CH-, and –CH2-), with the proviso that said compounds are not borrelidin (1), 12-desnitrile-12-carboxyl borrelidin (2), 10-desmethyl borrelidin (3), 11-epiborrelidin (4) or C14,C15-*cis* borrelidin analogue (5) as shown in Figure 1. In preferred embodiments:

(a). R_7, R_8 and R_9 are all CH_3.

(b). R_4 is CH_3 or COOH

(c). R_7, R_8 and R_9 are all CH_3.and R_4 is CH_3 or COOH

(d). R_1 is cyclobutane-1'-carboxylate

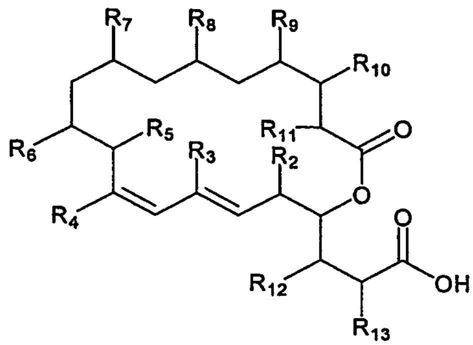
(e). R_1 is cyclobutane-1'-carboxylate and R_7, R_8 and R_9 are all CH_3.

(f). R_6, R_7, R_8 and R_9 are all CH_3, R_2 and R_{11} are H, R_5 and R_{10} are OH, R_4 is either CH_3, COOH or CN and R_1 is cyclopentane-1'-carboxylate or. cyclobutane-1'-carboxylate

(g). R_1 is cyclobutane-1'-carboxylate, R_7, R_8 and R_9 are all CH_3 and R_4 is CH_3 or COOH.

The present invention also provides compounds of formula 2 and pharmaceutically acceptable salts thereof, wherein:

Formula 2



R_2, R_3, R_6, R_7, R_8, R_9 or R_{11} are each independently H, OCH_3, CH_3 or CH_2CH_3; R_4 is CN, CO_2H, CHO, CH_3, $CONH_2$, CHNH, R_5, R_{10} are OH; or analogues differing from the corresponding "natural" compound in the oxidation state of one or more of the ketide units as shown in Figure 2 (i.e. selection of alternatives from the group: -CO-, -CH(OH)-, =CH-, and $-CH_2-$), and R_{12} and R_{13} are independently H or a C1-C4 alkyl group which may be optionally substituted with OH, F, Cl, SH) with the proviso that R_{12} and R_{13} are not simultaneously H.

In preferred embodiments:

(a). R_7, R_8 and R_9 are all CH_3.

(b). R_4 is CH_3 or COOH

(c). R_7, R_8 and R_9 are all CH_3.and R_4 is CH_3 or COOH

(d). R_{12} and R_{13} are independently CH_3 or H

(e). R_{12} and R_{13} are independently CH_3 or H and R_7, R_8 and R_9 are all CH_3

(f). R_6, R_7, R_8 and R_9 are all CH_3, R_2 and R_{11} are H, R_5 and R_{10} are OH, R_4 is either CH_3, COOH or CN and R_{12} and R_{13} are independently CH_3 or H

(g). R_6, R_7, R_8 and R_9 are all CH_3, R_2 and R_{11} are H, R_5 and R_{10} are OH, R_4 is either CH_3, COOH or CN and R_{12} and R_{13} are both CH_3

(h). R_{12} and R_{13} are independently CH_3 or H, R_7, R_8 and R_9 are all CH_3 and R_4 is CH_3 or COOH.

The compounds of the present invention may have tRNA synthetase-inhibitory activity (e.g. they may inhibit threonyl-, tyrosinyl-, or tryptophanyl-tRNA synthetase). They may display anti-microbial activity, including activity against intra- or extracellular parasites and organisms such as bacteria, spirochetes (e.g. Treponema), malaria, viruses and fungi. Additionally or alternatively they may have anti-proliferative activity against mammalian cells, and/or anti-angiogenic activity, either as a result of tRNA

synthetase inhibition, or through some other mode of action. This may make the compounds of the present invention particularly suitable as anti-cancer agents (e.g. agents for treatment of bowel cancer, prostate cancer or others), and may also provide application in treatment of other proliferative disorders, such as psoriasis, or conditions in which inappropiate vascularisation occurs, such as psoriasis, rheumatoid arthritis, atherosclerosis and diabetic retinopathy.

The compounds of the present invention may be formulated into pharmaceutically acceptable compositions, e.g. by admixture with a pharmaceutically acceptable excipient, carrier, buffer, stabiliser or other materials well known to those skilled in the art. Such compositions also fall within the scope of the present invention.

Such pharmaceutically acceptable materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material may depend on the route of administration, e.g. oral, intravenous, cutaneous or subcutaneous, nasal, intramuscular, intraperitoneal routes.

Pharmaceutical compositions for oral administration may be in tablet, capsule, powder or liquid form. A tablet may include a solid carrier such as gelatin or an adjuvant. Liquid pharmaceutical compositions generally include a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene glycol may be included.

For intravenous, cutaneous or subcutaneous injection, or injection at the site of affliction, the active ingredient will be in the form of a parenterally acceptable aqueous solution which is pyrogen-free and has suitable pH, isotonicity and stability. Those of relevant skill in the art are well able to prepare suitable solutions using, for example, isotonic vehicles such as Sodium Chloride Injection, Ringer's Injection, Lactated Ringer's Injection. Preservatives, stabilisers, buffers, antioxidants and/or other additives may be included, as required. Examples of the techniques and protocols mentioned above can be found in Remington's Pharmaceutical Sciences, 20th Edition, 2000, pub. Lippincott, Williams & Wilkins.

The invention further provides the compounds and compositions described above for use in a method of medical treatment. Also provided is the use of the compounds of the invention in the preparation of a medicament for the treatment of microbial conditions (including malaria), for the inhibition of angiogenesis, for the

treatment of proliferative disorders, or for the treatment of conditions characterised by inappropiate vascularisation, as decribed above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the structure of borrelidin and some related metabolites isolated from borrelidin producing organisms.

Figure 2 illustrates the incorporation patterns for ^{13}C stable isotope labelled extension substrates and the position of the *trans*-cyclopentane-1,2-dicarboxylic acid starter unit derived carbons.

Figure 3 illustrates the organisation of the borrelidin biosynthetic gene cluster. Restriction sites: B, *Bam*HI; Bc, *Bcl*I; E, *Eco*RI; X, *Xho*I.

Figure 4 illustrates a scheme showing the proposed biosynthetic pathway for the *trans*-cyclopentane-1,2-dicarboxylic acid starter unit.

Figure 5 illustrates the organisation of the borrelidin PKS and the biosynthesis of the pre-borrelidin molecule.

Figure 6 illustrates the proposed biosynthetic route for the introduction of the nitrile moiety at the C12 position of borrelidin.

Figure 7 illustrates the proposed structure of the molecule **6**.

Figure 8 illustrates the proposed structure of the molecules **7** & **8**.

Figure 9 illustrates the molecular characterisation of the 4-hydroxyphenylacetic acid catabolic pathway in *E. coli* W.

Figure 10 illustrates the structures of the molecules **18–20**

Figure 11 illustrates the structures of the molecules **21-26**

DETAILED DESCRIPTION OF THE INVENTION

A cosmid library of *S. parvulus* Tü4055 genomic DNA was constructed using fragments obtained from a partial digestion with *Sau3*AI that were cloned into pWE15 and introduced into *E. coli* cells using the Gigapack® III Gold Packaging Extract kit (Stratagene). A library of 3000 *E. coli* transformants was screened for homology using a labelled probe that was generated using the DIG DNA Labelling and Detection Kit (Roche). The probe used was a 1.7 kbp *Bgl*II-*Bam*HI fragment obtained from the gene that encodes module 6 of the third subunit of the oleandomycin PKS from *Streptomyces antibioticus* (Swan *et al.*, 1994).

Clones that gave a positive response were selected and cosmid DNA isolated. Cosmid DNA was digested with *Bam*HI and fragments less than 3 kbp in size were sub-cloned into pOJ260 (Bierman *et al.*, 1992). The plasmids were then used to transform *S. parvulus* Tü4055 protoplasts and resulting mutants were screened for the ability to produce borrelidin. Two mutants were identified as borrelidin non-producers, both of which were derived from plasmids that contained fragments of cosBor32A2. These two fragments were of 1.97 and 2.80 kbp in size, and were later identified as adjacent fragments encoding parts of the borrelidin PKS (*borA2* & *borA3*). Using cosBor32A2 as the probe, a second overlapping cosmid, cosBor19B9 was identified from the original library. These two cosmids are sufficient to cover the entire borrelidin biosynthetic gene cluster (see figure 3).

The complete nucleotide sequence of cosBor32A2 and cosBor19B9 was determined by shotgun sequencing of a *Sau*3AI-derived subclone library for each cosmid, consisting of 1.5-2.0 kbp fragments in pHSG397 (Takeshita *et al.*, 1987). Specific details are provided in example 3. The complete, overlapping nucleotide-coding sequence for cosBor32A2 and cosBor19B9 is presented as SEQ ID No.1. The region encoded by cosmid cosBor32A2 represents the sequence from nucleotide positions 0-40217 bp of SEQ ID No.1. The region encoded by cosmid cosBor19B9 overlaps this region by 4452 nucleotides, and corresponds to the nucleotide positions 35766-74787 bp of SEQ ID No.1. As described in more detail in the following text, we have performed gene inactivation experiments on many of the orfs identified to be encoded within SEQ ID No.1, and this leads us to identify the limits of the cluster. The borrelidin biosynthetic gene cluster is contained between nucleotide positions 7603 to 59966 of SEQ ID No.1 (*borB* to *borO,* which includes the *borA* region). Thus, these combined efforts have led us to the identification and sequencing of the DNA region encompassing the entire borrelidin biosynthetic gene cluster, and to the identification and description of the functional sequences encoded within this region.

PKS GENES

Encoded between positions 16184-50742 of SEQ ID No.1 are 6 orfs that display very high homology to the genes that encode the PKSs of known macrolide producing organisms. These genes are designated *borA1*, *borA2*, *borA3*, *borA4*, *borA5* and *borA6*, and encode the borrelidin PKS as was demonstrated above by disruption of a 1.97 kbp region within *borA2*. The six orfs are arranged in a head-to-tail manner and

each is terminated by an in-frame stop codon. The nucleotide sequence and corresponding polypeptide sequence details are shown below in Table 1:

Table 1

PKS encoding gene	Nucleotide position in SEQ ID No.1	Corresponding polypeptide sequence number
borA1	16184-18814	SEQ ID No.2
borA2	18875-23590	SEQ ID No.3
borA3	23686-34188	SEQ ID No.4
borA4	34185-39047	SEQ ID No.5
borA5	39122-45514	SEQ ID No.6
borA6	45514-50742	SEQ ID No.7

The gene *borA1* encodes the starter or loading module (SEQ ID No.1, position 16184–18814). The assignment of the start codon is not obvious for this open reading frame. The start codon given here is what we believe to be the true start codon, but there are at least another three possible start codons between the first and the beginning of the AT0 domain sequence and a person of skill in the art will appreciate that it may be possible to generate active protein using one of these alternative start codons. The start codon given here leaves a significant N-terminal tail of 321 amino acids preceding the AT0 domain. For comparison the N-terminal tail preceding the AT0 of the erythromycin loading module is 108 amino acids and that of the avermectin loading module is 28 amino acids. It is therefore possible that one of the other candidate start codons could be correct; the most likely of these are at positions 16298, 16607 and 16901 of SEQ ID No.1. The length of the N-terminal tail suggests it could possibly represent a catalytic activity, although it does not have any significant homology to other sequences in the databases. The nucleotide sequence position and the corresponding amino acid sequence for each of the functional domains within the starter module are identified below in Table 2:

Table 2

Domain in *borA1*	Bases in SEQ ID No.1	Amino acids in SEQ ID No.2
AT0	17147-18175	322-664
ACP0	18263-18472	694-763

The gene *borA2* encodes the first extension module (SEQ ID No.1, position 18875-23590). The nucleotide sequence position and the corresponding amino acid sequence for each of the functional domains within the first extension module are identified below in Table 3:

Table 3

Domain in *borA2*	Bases in SEQ ID No.1	Amino acids in SEQ ID No.3
KS1	18974-20251	34-459
AT1	20543-21529	557-885
KR1	22280-23011	1136-1379
ACP1	23129-23332	1419-1486

The gene *borA3* encodes the second and third extension modules (SEQ ID No.1, position 23686-34188). The nucleotide sequence position and the corresponding amino acid sequence for each of the functional domains within the second and third extension modules are identified below in Table 4:

Table 4

Domain in *borA3*	Bases in SEQ ID No.1	Amino acids in SEQ ID No.4
KS2	23785-25062	34-459
AT2	25360-26346	559-887
DH2	26392-26835	903-1050
KR2	27745-28476	1354-1597
ACP2	28567-28767	1628-1694
KS3	28855-30132	1724-2149
AT3	30418-31413	2245-2576
DH3	31462-31887	2593-2734
KR3	32863-33606	3060-3307
ACP3	33703-33903	3340-3406

The gene *borA4* encodes the fourth extension module (SEQ ID No.1, position 34185-39047). The nucleotide sequence position and the corresponding amino acid sequence for each of the functional domains within the fourth extension module are identified below in Table 5:

Table 5

Domain in *borA4*	Bases in SEQ ID No.1	Amino acids in SEQ ID No.5
KS4	34284-35561	34-459
AT4	35847-36842	555-886
KR4	37719-38453	1179-1423
ACP4	38559-38759	1459-1525

The gene *borA5* encodes the fifth extension module (SEQ ID No.1, position 39122-45514). The nucleotide sequence position and the corresponding amino acid sequence for each of the functional domains within the fifth extension module are identified below in Table 6:

Table 6

Domain in *borA5*	Bases in SEQ ID No.1	Amino acids in SEQ ID No.6
KS5	39221-40492	34-457
AT5	40778-41785	553-888
DH5	41834-42259	905-1046
ER5	43322-44191	1401-1690
KR5	44207-44947	1696-1942
ACP5	45044-45244	1975-2041

The gene *borA6* encodes the sixth extension module and the chain terminating thioesterase (SEQ ID No.1, position 45514-50742). The nucleotide sequence position and the corresponding amino acid sequence for each of the functional domains within the sixth extension module are identified below in Table 7:

Table 7

Domain in *borA6*	Bases in SEQ ID No.1	Amino acids in SEQ ID No.7
KS6	45622-46884	37-457
AT6	47176-48162	555-883
KR6	48814-49518	1101-1335
ACP6	49624-49824	1371-1437
TE	49894-50637	1461-1708

The identification of functional domains and their boundaries as described in the aforementioned are determined based on the similarities to the conserved amino acid sequences of other modular PKSs such as those for the rapamycin (Schwecke *et al.*, 1995; Aparicio *et al.*, 1996) and erythromycin (Cortés *et al.*, 1990) biosynthesis. The limits of the catalytic domains are established on the basis of homology to other PKS clusters and the chosen point at which a domain starts or finishes is not absolutely defined, but selected based on the aforementioned considerations. In the case of the β-keto processing domains it is least obvious, as there is typically a large region not assigned to a functional domain that precedes the KR domain. This region may be structurally important, or required for stability of the PKS dimer. An unusual characteristic of the borrelidin PKS is that all of the individual enzymatic domains appear to be catalytically competent based on their oligonucleotide/amino acid sequence, and are all necessary in order to provide the β-keto processing required to produce the functional groups observed in borrelidin. This is rather unusual as the majority of modular PKS sequences so far reported contain one or more inactive domains, an exception being for example the spinosyn PKS (Waldron *et al.*, 2001; US 6,274,50).

One skilled in the art is familiar with the degeneracy of the genetic code, therefore, the skilled artisan can modify the specific DNA sequences provided by this disclosure to provide proteins having the same or altered or improved characteristics compared to those polypeptides specifically provided herein. One skilled in the art can also modify the DNA sequences to express an identical polypeptide to those provided, albeit expressed at higher levels. Furthermore, one skilled in the art is familiar with means to prepare synthetically, either partially or in whole, DNA sequences which would be useful in preparing recombinant DNA vectors or coding sequences which are encompassed by the current invention. Additionally, recombinant means for modifying the DNA sequences provided may include for example site-directed deletion or site-directed mutagenesis. These techniques are well known to those skilled in the art and need no further explanation here. Consequently, as used herein, DNA which is isolated from natural sources, prepared synthetically or semi-synthetically, or which is modified by recombinant DNA methods, is within the scope of the present invention.

Likewise, those skilled in the art will recognize that the polypeptides of the invention may be expressed recombinantly. Alternatively, those polypeptides may be synthesised either in whole or in part, by conventional known non-recombinant

techniques; for example, solid phase synthesis. Thus, the present invention should not be construed as necessarily limited to any specific vector constructions or means for production of the specific biosynthetic cluster molecules including the polyketide synthase molecules exemplified.

The loading module of the borrelidin PKS exists as a discrete protein. This is rather unusual as the majority of loading modules are found on the same protein as the first extension module. Exceptions to this include, for example, the nystatin (Brautaset *et al.*, 2000) and amphotericin (Caffrey *et al.*, 2001) PKSs. The loading module, which consists of an AT-ACP didomain, is similar to the broad specificity loading module of the avermectin PKS, which accept a number of alternative starter acids, and are of use in generating libraries of novel polyketides (Marsden *et al.*, 1998; Pacey *et al.*, 1998). The AT domain of the borrelidin PKS loading module diverges from the vast majority of AT domains as the active site serine residue is replaced with a cysteine such that the active site motif is GXCXG (specifically GHCYG). In most available type-I PKS AT domain sequences, the conserved active site motif is GXSXG; the same motif is observed in lipases, fatty acid synthases and most thioesterases. The nucleophilic serine is substituted by cysteine in two NRPS thioesterase domains, specifically the synthetases responsible for the production of mycobactin and pyochelin (Shaw-Reid *et al.*, 1999). A GXCXG motif is also observed in a thioesterase-like domain of ORF1 in the bialaphos cluster (Raibaud *et al.*, 1991). It has been suggested that since it is not possible to move between the two types of serine codons by a single base change, active sites containing an essential serine residue may lie on two lines of descent from an ancient ancestral enzyme that had a cysteine instead of a serine in its active site (Brenner, 1988). The presence of enzymes containing cysteine in the active site may support this view. It may alternatively be the case that cysteine arises in these active sites because it is possible to move from one type of serine codon to the other via a cysteine which would remain active.

The AT domains of PKSs select a particular carboxylic acid unit as substrate. This selectivity has been shown to correlate with certain motif signatures within the AT domain (Reeves *et al.*, 2001; WO 02/14482). The borrelidin loading module AT domain motif differs from any described so far, which is not surprising as this AT domain is the first to be sequenced that selects an alicyclic dicarboxylic acid. The AT domains for the borrelidin PKS extension modules display the expected active site motif GXSXG, and also each contain the expected motifs for the selection of malonyl-CoA or

methylmalonyl-CoA (Reeves *et al.*, 2001; WO 02/14482). The malonyl-CoA selective AT domains (AT1, AT2 and AT6) show very high similarity to one another, both at the protein and at the DNA level. The same is true for the methylmalonyl-CoA selective AT domains (AT3, AT4 and AT5); two of these AT domains (AT3 and AT4) have identical amino acid sequences throughout the conserved region. The high similarity of AT5 to AT3 and AT4 is evidence that the extender unit selected in module 5 is methylmalonyl-CoA, and that the borrelidin C12-methyl group thus incorporated is subsequently modified to a nitrile function after incorporation into the PKS.

To demonstrate that we can alter the PKS derived structure of borrelidin, the AT domain of module 4 (the AT domain encoded by *borA4*) is replaced by the AT domain of module 2 of the rapamycin PKS (rapAT2) using a replacement strategy (see example 6). This gives strain *S. parvulus* Tü4055/467. Upon fermentation and LCMS analysis of culture extracts of this mutant, it can be determined that some borrelidin is produced and a new, more polar compound is also observed with a *m/z* value 14 units lower than borrelidin. This is consistent with incorporation of a malonate rather that a methylmalonate extender unit by module 4 of the PKS to produce 10-desmethyl borrelidin **3**.

In addition to production by domain swapping methods, **3** is also generated by introducing specific mutations into the module 4 AT domain selectivity motif (Reeves *et al.*, 2001; WO 02/14482) (see example 7). Such a change affects the selectivity of the AT domain such that it selects a substrate molecule of malonyl-CoA preferentially over methylmalonyl-CoA. Thus, the amino acid motif YASH at positions 739 to 742 of SEQ ID No.5 is mutated to HAFH to give strain *S. parvulus* Tü4055/472. Upon fermentation and LCMS analysis of culture extracts of this mutant it is determined that borrelidin is produced in addition to a new, more polar compound with a *m/z* value 14 units lower than borrelidin. This new compound is identical to that described above and thus is consistent with incorporation of a malonate rather that a methylmalonate extender unit by module 4 of the PKS to produce **3**.

These results clearly indicate that the borrelidin PKS is amenable to genetic manipulation and to the exchange of native sequence for that of a heterologous strain. It is clear to one skilled in the art that the biosynthetic engineering, by the methods described above, of the borrelidin PKS will lead to the production of novel borrelidin-like molecules.

The borrelidin loading module is of interest due to the unique structure of its cognate substrate. To examine its potential use in other systems, the loading module native to the erythromycin PKS is replaced with the borrelidin loading module in *Saccharopolyspora erythraea*; this experiment is analogous to those done previously with the avermectin loading module (WO 98/01546; Marsden *et al.*, 1998). We anticipate that the new strain is capable of producing novel erythromycin like molecules in which the C13-ethyl group is replaced with an exogenously supplied racemic *trans*-cyclopentane-1,2-dicarboxylic acid moiety. The methodology used to perform this experiment is similar to that described in WO 98/01546, but the transformation is performed using a mutant *Saccharopolyspora erythraea* DM (Gaisser *et al.*, 2000) which accumulates the aglycone product erythronolide B rather than the fully processed macrolide, as well as using *S. erythraea* WT. This experiment is described in example 8.

It is not evident from SEQ ID No.1, which of four candidate start codons is correct for *borA1*. The four most obvious candidate start codons are at nucleotides 16184, 16298, 16607 and 16901 of SEQ ID No.1. The earliest of these possible start codons was used in giving the amino acid sequence for SEQ ID No.2. A pile-up of this loading module with the erythromycin and avermectin loading modules indicates that the AT0 domain starts at position 321 of SEQ ID No.2, and that there is a long N-terminal tail. No significant homology is found for the first 298 amino acids of *borA1*. The borrelidin loading module is encoded by a discrete orf, and in order to retain this architecture the splice site chosen for joining the borrelidin PKS loading module sequence to the erythromycin PKS loading module sequence is at the beginning of the homologous region of the KS1 domain of *borA2*, at amino acids 42-44 of SEQ ID No.3. This approach maintains the putative docking regions at the end of BorA1 and start of BorA2 that are believed to be essential for the production of a functional PKS assembly. To maintain the continuity of this experiment this loading module is fused to the equivalent point at the beginning of the KS1 domain of *eryA1*. The resulting mutants *S. erythraea* DM/CJM400-403 are fermented and analysed by negative ion LCMS using standard protocols. This analysis clearly indicates the presence of a new compound **6** with *m/z* = 485.3 as expected (figure 7). It is clear to one skilled in the art that the products of these experiments could be biotransformed using an appropriate strain such as *S. erythraea* JC2 (Rowe *et al.*, 1998) to provide novel, biologically active erythromycin analogues. It is additionally clear to one skilled in the art that the

borrelidin loading module has utility for the biosynthetic engineering of other PKSs (i.e. not the borrelidin PKS) to produce further novel polyketides bearing a *trans*-cyclopentane-1,2-dicarboxylic acid moiety. It is also clear that the diversity of products arising from hybrid PKSs derived from the borrelidin loading module may be further enhanced through the exogenous feeding of carboxylic acids other than the cognate substrate.

The most striking feature of the borrelidin PKS is the clear divergence from the normal co-linear, processive mode of operation for type-I modular PKSs. Borrelidin is a nonaketide (expected: one loading plus eight extension steps), but only seven modules (one loading and six extension modules) are present in the cluster. Analysis of the PKS domains with respect to the chemical structure of borrelidin correlates with the fifth extension module (BorA5) being used iteratively for three rounds of chain elongation as shown in figure 5. Thus, the fifth, sixth and seventh rounds of chain elongation occur on BorA5 with the incorporation of three methylmalonyl-CoA extension units, and with full reductive processing of the β-keto groups to methylene moieties. As described *supra*, the divergence from co-linear operation for modular PKSs is unusual and limited to a few examples. The present example is interesting as it occurs on a module that reduces the β-keto group fully to a methylene moiety and which is followed by an inter- rather than intra-protein transfer of the growing chain. This is also the case for the two known examples of erroneous iterative use of type-I modules by the erythromycin (Wilkinson *et al.*, 2000) and epothilone (Hardt *et al.*, 2001) PKSs. It is noteworthy that this full reduction makes these modules functionally equivalent to fatty acid synthase (FAS). The type-I PKS modules that can operate iteratively may have retained FAS like activity.

Although it appears that BorA5 is used iteratively (three times), two other possible scenarios may explain borrelidin biosynthesis given the genes present in the borrelidin biosynthetic cluster. Firstly, two modules may be 'missing' from the cluster, but could be present at some other location in the genome. However, in the majority of cases investigated, the genes required for biosynthesis of secondary metabolites in actinomycetes are clustered in a single locus. The second possibility is that three separate BorA5 dimers assemble, and that each catalyses a round of chain elongation; thus the process would be processive. However, this scenario requires that three times the amount of BorA5 is produced with respect to the other PKS proteins, but the organisation of the borrelidin gene cluster does not indicate that the regulation of *borA5*

differs from that of any of the other PKS genes. In addition, this scenario does not fit with the common thinking as to the roles of inter-protein docking domains, which suggests that there is a specific recognition between the N- and C-terminal ends of the proteins of the biosynthetic complex that need to interact, enabling specific binding between modules encoded on different proteins (Ranganathan *et al.*, 1999; Wu *et al.*, 2001; Broadhurst *et al.*, 2003).

To address the issues described above, the two proteins encoded by *borA4* and *borA5* were fused after manipulation at the genetic level to provide strain *S. parvulus* Tü4055/*borA4-A5* (see example 9), and separately the two proteins encoded by *borA5* and *borA6* were fused in an analogous manner to provide strain *S. parvulus* Tü4055/*borA5-A6* (see example 10). Additionally, a double mutant was generated in which the above described fusions were combined to generate a strain in which *borA4*, *borA5* and *borA6* were fused to generate strain *S. parvulus* Tü4055/*borA4-A5-A6* (see example 11). Therefore, the new, fused, bi- and tri-modular genes make it impossible to assemble three separate molecules of BorA5, or for another protein(s) encoded by a gene(s) remote from the borrelidin cluster to act in tandem with BorA5. Upon fermentation of strains *S. parvulus* Tü4055/*borA4-A5*, /*borA5-A6*, and /*borA4-A5-A6* followed by extraction and analysis, the production of borrelidin was verified at a reduced but significant level (21±4%, 27±4% and 18±5% respectively) when compared to the WT strain. Thus, the production of borrelidin by these mutants indicates that module 5 of the fused BorA4-A5 or BorA5-A6 operates in an iterative manner. Since the priority filing of this application, these limited data have been published (Olano *et al.*, 2003).

The ability of BorA5 to operate iteratively has great potential for the engineering of heterologous PKSs to provide macrolactones with expanded ring sizes. To examine this possibility BorA5 is swapped into the erythromycin PKS in place of module 4 of DEBS2. This is done by replacement of the appropriate gene fragment in both the erythromycin producer *S. erythraea* WT and *S. erythraea* DM. This experiment is chosen as both modules recruit methylmalonyl-CoA extender units and process the β-keto functions formed through to methylene groups. In addition, the stereochemistry of the resulting methyl group in the polyketide chain is the same in both cases. Of most significance is the fact that module 4 of DEBS2 is known to perform erroneous iterative rounds of chain elongation (Wilkinson *et al.*, 2000), indicating that such a process can indeed occur at this location within the PKS and give rise to products that can be fully

processed by DEBS3, making it an attractive target to introduce specific iterative use of a heterologous module to make 16- and 18-membered macrolides.

Briefly, the region of DNA encoding *borA5* is swapped for that encoded by module 4 of *eryA2*, which encodes the C-terminal portion of DEBS2 of the erythromycin PKS (see example 12). The resulting mutant *S. erythraea* DM/421 is grown and extracted as for the production of metabolites by *S. erythraea* strains (Wilkinson *et al.*, 2000) and then analysed by LCMS. Two new significant compounds, which are less polar than erythronolide B, are observed. These have an *m/z* of 435.5 (**7**, [MNa^+]) and 477.5 (**8**, [MNa^+]) respectively, which is consistent with the production of two new ring expanded erythronolide B analogues (figure 8). Compound **7** with *m/z* = 435.5 is consistent with the presence of the 16-membered ring-expanded erythronolide B related macrolide reported previously as a minor component of *S. erythraea* WT fermentations (Wilkinson *et al.*, 2000). It is clear to one skilled in the art that such new products can be converted to antibacterial molecules by biotransformation with an appropriate organism. It is also clear to one skilled in the art, that the inclusion of such a module into other positions of the erythromycin PKS, or into other PKSs, may allow the production of novel, ring expanded polyketides in a similar manner. In addition, it is possible to perform this experiment by swapping only the region of the DEBS module 4 from the start of the conserved region of the KS4 to the end of the ACP4 domain; this arrangement retains the C- and N-terminal regions at the end of DEBS2 and DEBS3 respectively, to ensure the mutual recognition and docking of these proteins.

NON-PKS GENES

Both upstream and downstream of the PKS encoding genes are other orfs involved in the biosynthesis of borrelidin. An orf is designated as consisting of at least 100 contiguous nucleotides, that begins with an appropriate start codon and finishes with an appropriate stop codon, and which has an appropriate codon bias for protein-coding regions of an organism whose DNA is rich in the nucleotides guanine and cytosine. In the DNA sequence both upstream and downstream of the borrelidin PKS genes (*borA1-borA6*) there are a number of orfs that could be identified by comparison to other sequences in the NCBI database (see figure 3). The nucleotide sequence details of these orfs are given below in Table 8:

Table 8

Gene	Bases in SEQ ID No.1	Corresponding polypeptide sequence number
borB	7603-8397c	SEQ ID No.8
borC	8397-9194c	SEQ ID No.9
borD	9244-9996c	SEQ ID No.10
borE	9993-11165c	SEQ ID No.11
borF	11162-11980c	SEQ ID No.12
borG	11992-13611c	SEQ ID No.13
borH	13608-15599c*	SEQ ID No.14
borI	50739*-52019	SEQ ID No.15
borJ	52113-53477	SEQ ID No.16
borK	53486-54466	SEQ ID No.17
borL	54506-56176	SEQ ID No.18
borM	56181*-57098	SEQ ID No.19
borN	57112-57858	SEQ ID No.20
borO	57939-59966	SEQ ID No.21
orfB1	2-313	SEQ ID No.22
orfB2	501*-3107	SEQ ID No.23
orfB3	3172-3810c	SEQ ID No.24
orfB4	3935-4924c	SEQ ID No.25
orfB5	5123-5953	SEQ ID No.26
orfB6	5961-6518*c	SEQ ID No.27
orfB7	6564*-7538	SEQ ID No.28
orfB8	60153-60533c	SEQ ID No.29
orfB9	60620-61003	SEQ ID No.30
orfB10	61188*-61436	SEQ ID No.31
orfB11	61526-61738	SEQ ID No.32
orfB12	61767-62285c	SEQ ID No.33
orfB13a	62750-63067c	SEQ ID No.34
orfB13b	62586-62858c	SEQ ID No.113
orfB14	63155-65071c	SEQ ID No.35
orfB15	65374-65871	SEQ ID No.36
orfB16	65942-68305c*	SEQ ID No.37

Gene	Bases in SEQ ID No.1	Corresponding polypeptide sequence number
orfB17	68290-68910c*	SEQ ID No.38
orfB18	69681-70436	SEQ ID No.39
orfB19	70445-71848	SEQ ID No.40
orfB20	71851-72957	SEQ ID No.41
orfB21	73037-73942	SEQ ID No.42
orfB22	73995-74534c	SEQ ID No.43

[Note 1: c indicates that the gene is encoded by the complement DNA strand; Note 2: for each open reading frame given above, the longest probable open reading frame is described. It is sometimes the case that more than one potential candidate start codon can been identified. One skilled in the art will recognise this and be able to identify alternative possible start codons. We have indicated those genes which have more than one possible start codon with a '*' symbol. Throughout we have indicated what we believe to be the start codon, however, a person of skill in the art will appreciate that it may be possible to generate active protein using an alternative start codon, proteins generated using these alternative start codons are also considered within the scope of the present invention. Note 3 the SEQ ID NO: for *orfB13b* was originally designated SEQ ID NO: 34 but for clarity a separate sequence and SEQ ID NO has been assigned.]

Potential functions of the predicted polypeptides (SEQ ID No.7 to 43) were obtained from the NCBI database using a BLAST search. The best matches obtained from these searches are described below in Table 9:

Table 9

Gene	Significant protein match	Score	Accession GenBank	Proposed function
orfB1	hypothetical protein, no full length hits, high GC codon preference			unknown
orfB2	SCM2.07, hypothetical protein (*S. coelicolor*)	998	NP_625154	unknown
orfB3	SCF76.07, hypothetical protein, (*S. coelicolor*)	359	NP_624786	unknown

Gene	Significant protein match	Score	Accession GenBank	Proposed function
orfB4	SCF76.06, araC family transcriptional regulator (*S. coelicolor*)	412	NP_624785	unknown
orfB5	SCF76.05c, non-heme chloroperoxidase (*S. coelicolor*)	495	NP_624784	non-heme chloroperoxidase
orfB6	SCF76.09, hypothetical protein (*S. coelicolor*)	159	NP_624788	unknown
orfB7	SCF76.08c, hypothetical protein (*S. coelicolor*)	473	NP_624787	unknown
borB	PteH, polyene macrolide thioesterase (*S. avermitilis*)	244	BAB69315	type II thioesterase
borC	XF1726, 2,5-dichloro-2,5-cyclohexadiene-1,4,-diol dehydrogenase (*Xylella fastidiosa* strain 9a5c)e	160	NP_299015	dehydrogenase
borD	FabG, 3-oxoacyl-ACP reductase precursor, (*Plasmodium falciparum*)	124	AAK83686	3-oxoacyl-ACP reductase
borE	FN1586, O-succinylbenzoyl-CoA synthase, (*Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586)	88	NP_602402	cyclase (member of enolase superfamily)
borF	putative lysophospholipase homologue, (*Arabidopsis thaliana)*	57	NP_565066	unknown
borG	MTH1444, acetolactate synthase, large subunit, (*Methanothermobacter thermautotrophicus*)	120	NP_276558	Unknown

Gene	Significant protein match	Score	Accession GenBank	Proposed function
borH	PA3592, conserved hypothetical protein, (*Pseudomonas aeruginosa*)	116	NP_252282	unknown
borI	TylH1, cytochrome P450, (*Streptomyces fradiae*)	285	AAD12167	cytochrome P450 oxidase
borJ	BioA, DAPA aminotransferase, (*Kurthia* sp. 538-KA26)	346	BAB39453	amino transferase
borK	Adh1, alcohol dehydrogenase, (*Aquifex aeolicus*)	191	NP_213938	NAD/quinone oxidoreductase
borL	putative auxin-regulated protein GH3, (*Arabidopsis thaliana*)	92	NP_176159	unknown
borM	SCL6.10, hypothetical protein similar to putative F420-dependent dehydrogenase (*S. coelicolor*),	108	CAB76875	F420 dependent dehydrogenase
borN	SC1C2.27, hypothetical protein, 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase superfamily (*S. coelicolor)*	215	NP_629680	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase
borO	ThrS, threonyl-tRNA synthetase (*Mycobacterium leprae*)	627	NP_301410	threonyl-tRNA synthetase, self resistance gene
orfB8	conserved hypothetical protein (*Methanosarcina acetivorans* str. C2A). (Pfam pulls out weak MarR family)	37	NP_617908	possible regulator

Gene	Significant protein match	Score	Accession GenBank	Proposed function
orfB9	putative anti-sigma factor antagonist (*Streptomyces coelicolor*)	113	NP_631789	anti-sigma factor antagonist
orfB10	conserved hypothetical protein (*S. coelicolor*)	95	NP_631790	unknown
orfB11	hypothetical protein, no full length hits, high GC codon preference			unknown
orfB12	putative regulator (*S. coelicolor*)	92	NP_631494	regulator (of a two component system, maybe membrane sensor)
orfB13a *orfB13b*	putative acetyltransferase (*S. coelicolor*); putative acetyltransferase (*S. coelicolor*)	58 100	NP_625155 NP_625155	tentative assignment of acetyltransferase in two frames, or sequencing error and should be in a single frame
orfB14	putative lipoprotein (*S. coelicolor*)	386	NP_631245	unknown
orfB15	hypothetical protein (*S. coelicolor*)	41	NP_631424	unknown
orfB16	putative formate dehydrognease (*S. coelicolor*) (Pfam matches to molybdopterin oxidoreductase/ formate dehydrogenase alpha subunit)	915	NP_626265	oxidoreductase

Gene	Significant protein match	Score	Accession GenBank	Proposed function
orfB17	conserved hypothetical protein, *S. coelicolor* SCBAC25F8.16	175	NP_631569	unknown
orfB18	product unknown (*Streptomyces aureofaciens*)	396	AAD23399	unknown
orfB19	putative aldehyde dehydrogenase (*S. aureofaciens*)	635	AAD23400	aldehyde dehydrogenase
orfB20	putative alcohol dehydrogenase (*S. coelicolor*)	450	NP_630527	alcohol dehydrogenase
orfB21	hypothetical protein (*S. coelicolor*)	395	NP_630528	unknown
orfB22	putative calcium binding protein (*S. coelicolor*)	160	NP_631687	calcium binding protein

Analysis of the functions of the putative gene products indicates that the genes *borB* to *borO* most probably form the boundaries of the borrelidin biosynthetic cluster. Evidence to support this came from the disruption of *borB2*, which produced borrelidin at levels indistinguishable from the wild type parental strain. In addition, *borB3* to *borB7* have homologues in the *Streptomyces coelicolor* A3(2) genome encoded on cosmid SCF76; the same orfs are present, but in a different order. The orfs *borB8* to *borB10* are arranged identically to homologues in the *S. coelicolor* A3(2) cosmid SC5E3. The orfs *borB18* to *borB21* have homologues that are arranged similarly in the *S. coelicolor* A3(2) cosmid SC1A2. The orf *borB13* contains a frame-shift and thus any gene product would most probably be inactive. In addition, no function can be readily deduced for the products of these orfs during borrelidin biosynthesis.

<u>Starter unit biosynthesis genes</u>

In order to identify the genes that are involved in the biosynthesis of the *trans*-cyclopentane-1,2-dicarboxylic starter unit, each of the genes *borB* to *borN* was disrupted (e.g. see examples 13-25). This was done in a manner designed to minimise the possibility of polar effects, which was verified by successful *in trans*

complementation with a full-length copy of the disrupted gene under the control of the ermE* promoter, which gave back approximately WT levels of borrelidin production in each case.

Each of the disrupted mutants was grown in triplicate as described in example 1, and borrelidin production assessed. Alongside these, each mutant was grown in triplicate and supplemented, after 24 hours, with exogenous starter acid to a final concentration of 1 mM, and borrelidin production assessed. Extraction and analysis for borrelidin provided the data that are described below in Table 10:

Table 10

Borrelidin biosynthetic gene disrupted	Borrelidin production without feeding (%relative to WT)	Borrelidin production with feeding (%relative to unfed WT)
Wild type (control)	100±16, (100±2)	363±65, (269±49)
borB	75±11, (43±20)	172±51
borC	0, (10±3)	933±42
borD	7±1, (0)	75±15
borE	2±1	122±23
borF	3±2	201±52
borG	11±1, (32±3)	1532±142
borH	17±2, (23±13)	203±40
borI	0, (0)	0, (0)
borJ	0, (0)	0, (0)
borK	0, (6±1)	319±54, (464±18)
borL	0, (0)	408±70, (399±69)
borM	0, (6±3)	461±29, (553±66)
borN	25±9, (34±3)	68±12, (46±9)
borO	N/A	N/A

[Note 1: The values given in brackets indicate where repeat runs of some experiments were performed; Note 2: N/A = not applicable.]

Based on the data in table 10, it is clear to one skilled in the art that the gene products BorC-F and K-M are essential or very important for the biosynthesis of *trans*-cyclopentane-1,2-dicarboxylic acid, as these mutants produced no or very low levels of borrelidin without the addition of exogenous starter acid, whereupon they produced borrelidin at levels approaching, or better than, that of the WT organism. In addition the

gene products BorG, H, and N appear to be involved in, but not essential for, the biosynthesis of the starter unit, as they produced significantly lower levels of borrelidin unless exogenous starter acid was added, whereupon they produced borrelidin at levels approaching or better than that of the WT organism; this was particularly notable in the case of the *borG*⁻ mutant.

The normal metabolic function of BorN homologues is the production of 2-oxohepta-3-ene-1,7-dioate **10**, a key step in the catabolism of tyrosine *via* 4-hydroxyphenyl acetic acid **9** (figure 9) (Prieto *et al.*, 1996). Therefore, **10** may be an intermediate in the biosynthetic pathway to *trans*-cyclopentane-1,2-dicarboxylic. The ability of the mutant disrupted in *borN* to produce borrelidin, albeit at a reduced level, most probably lies in the presence of a homologue elsewhere in the genome utilised in the catabolism of tyrosine during primary metabolism.

The intermediate **10** contains all the required functionality for the eventual formation of *trans*-cyclopentane-1,2-dicarboxylic acid. The most probable next step of the biosynthesis is the reduction of the 3-ene position in a reaction similar to that catalysed by an enoyl reductase. Potential enzymes responsible for this step are BorC, BorD, BorK or BorM; these enzymes are all involved in borrelidin starter unit biosynthesis as seen from the data in table 10. The resulting 2-oxohepta-1,7-dioate **11** is one possible substrate for cyclisation through formation of a new C-C bond between C6 and C2. Another possible substrate for this cyclisation would be 2-hydroxyhepta-1,7-dioate **12** or some activated form thereof. This would presumably be formed from **11** by the action of an oxidoreductase such as BorC, BorD or BorM.

The key cyclisation step is most probably catalysed by BorE, which displays similarity to O-succinylbenzoyl-CoA synthase and chloromuconate cycloisomerase. These enzymes belong to the enolase super-family, the members of which share the common ability to stabilise the formation of an anion on the carbon atom adjacent to a carboxylate group (Schmidt. *et al.*, 2001). It is further notable that the substrate for muconate cycloisomerase is a hexa-1,6-dioate, which is similar in gross structure to **11** and **12**. Abstraction of a proton and formation of an anion equivalent at C6 of **11** or **12** (or an activated form thereof, e.g. **13**) with subsequent cyclisation to C2 provides the correctly substituted cyclopentane ring structure, although the intermediacy of **11** as substrate would require some further processing of the substituted cyclopentane, most probably via elimination of water to give the symmetric cyclopent-1-ene-1,2-dicarboxylic acid, or possibly the Δ^1-unsaturated compound, cyclopent-1-ene-1,2-

dicarboxylic acid. However, the feeding of cyclopent-1-ene-1,2-dicarboxylic acid, or ethyl esters thereof, to *S. parvulus* Tü4055 strains disrupted in any of *borC-E*, or to WT strains, did not produce any borrelidin, or did not produce borrelidin in any increased amount when compared to the unfed controls. These data indicate that this compound is probably not an intermediate in starter unit biosynthesis, and that the substrate of BorE is possibly the 2-hydroxyhepta-1,7-dioate **12**, or an activated form thereof (e.g. **13**). A putative pathway for the biosynthetic pathway to *trans*-cyclopentane-1,2-dicarboxylic acid is shown in figure 4.

The combined, specific genes required for the biosynthetic steps to *trans*-cyclopentane-1,2-dicarboxylic acid are not clear, but probably are encoded by some combination of *borC-H*, *borK*, *borM* and *borN*. The lack of certain homologues of genes that are involved in the catabolism of 4-hydroxyphenyl acetic acid **9**, and which would act prior to BorN in the pathway, is most probably an indication that primary metabolic genes perform these tasks. The addition of exogenous *trans*-cylopentane-1,2-dicarboxylic acid to *S. parvulus* Tü4055 and related strains increases the titre of borrelidin in the order of 2- to 3-fold under our conditions, indicating that the biosynthesis of starter acid is a limiting factor in borrelidin biosynthesis. These data are consistent with primary metabolic degradation of tyrosine being the source of *trans*-cyclopentane-1,2-dicarboxylic acid.

In an attempt to further clarify which genes may be specifically responsible for biosynthesis of the starter unit, a number of co-culture experiments were performed with combinations of the different mutants – these require the knowledge that the gene products of borI and borJ are specifically involved in the formation of the C12-nitrile moiety, which is clarified by the data given in the following section below in combination with the data from table 10. In summary, the co-culture of mutants *borE⁻* & *borD⁻*, and of *borE⁻* & *borM⁻* failed to produce any borrelidin whereas the co-culture of mutants *borM⁻* & *borI⁻*, and *borM⁻* & *borK⁻* produced borrelidin at approximately WT levels. These data, in combination with that in table 10, and below, clearly indicate that *borD*, *borE* and *borM* are involved in starter unit biosynthesis, whereas *borI*, and possibly *borK*, are involved in the formation of the nitrile moiety at C12 of borrelidin.

It is clear from the data in table 10 that exogenous addition of *trans*-cyclopentane-1,2-dicarboxylic acid is sufficient to re-establish approximately WT levels, or better, of borrelidin production in mutants where genes that are involved in starter unit biosynthesis have been disrupted. These data indicate that there is no problem

with the active uptake of added carboxylic acid by *S. parvulus* Tü4055, and that an activity is present which is capable of converting the carboxylic acid to a CoA thioester equivalent. Thus,. given the known technologies of mutasynthesis, it is obvious to one skilled in the art that the addition of exogenous carboxylic acids to one of the aforementioned mutants, for example the *borE*⁻ strain *S. parvulus* Tü4055/*borE:aac3(IV)* described in example 16, may lead to the production of borrelidin analogues in which the starter unit carboxylic acid moiety is replaced with a moiety derived from the exogenously added carboxylic acid.

To examine this possibility, strain *S. parvulus* Tü4055/*borE:aac3(IV)* was fed with a *trans*-cyclobutane-1,2-dicarboxylic acid according to the protocol described in example 1 and then analysed as described in example 4. The structure **18**, described in figure 10, shows the new borrelidin structure obtained from feeding this carboxylic acid; this compound **18** displayed the anticipated UV chromophore for borrelidin but eluted at an earlier retention time and displayed the expected mass by LCMS (m/z = 474.3 $[M-H]^-$XX). Verification of this methodology was provided by the production, isolation and characterisation of **18** (example 33). (RS)-2It is clear to one skilled in the art that other carboxylic acids could also be used in similar feeding experiments to provide further new borrelidin analogues. Although it is possible that not all carboxylic acids would be incorporated using the exact methodology described herein, a person of skill in the art is aware of a number of available methods to enhance the incorporation of fed starter units.

In addition to the use of the strain deleted in *borE*, it was observed (see table 10) that the strain *S. parvulus* Tü4055/*borG:aac3(IV)*, in which *borG* has been disrupted, when fed with the natural starter unit of the bor PKS, *trans*-cyclopentane-1,2-dicarboxylic acid, produced borrelidin at titres significantly higher than those seen when the the wild-type organism was fed (4-fold increase) or unfed (15-fold increase). To examine this further, this experiment was repeated using both the natural and an unnatural starter acid as exogenous substrates, fed, in parallel, to wildtype, the *borE* mutant and the *borG* mutant. The resulting data are described in table 11.

Table 11:

	Unfed	Fed with 1 mM cyclopentane *trans*- 1,2-dicarboxylic acid	Fed with 1 mM cyclobutane trans-1,2-dicarboxylic acid
S. parvulus Tü4055	2.3 mg/l	6.6 mg/l	-
S. parvulus Tü4055/*borE:aac3(IV)*	0	4.7 mg/l	2.2 mg/l
S. parvulus Tü4055/*borG:aac3(IV)*	0	88.9 mg/l	43.0 mg/l

As one can see from table 11, using *S. parvulus* Tü4055/*borG:aac3(IV)* instead of *S. parvulus* Tü4055/*borE:aac3(IV)* for mutasynthesis increases the titre approximately 19-fold, and that *S. parvulus* Tü4055/*borG:aac3(IV)* fed with the natural starter acid produces 38-fold more borrelidin A than wild type alone, or 13 fold more borrelidin A than the wild type strain fed with the same amount of cyclopentane *trans*-1,2-dicarboxylic acid. These data clearly indicate that the use of strain *S. parvulus* Tü4055/*borG:aac3(IV)* for mutasynthesis experiments is beneficial for the production of improved titres of borrelidin analogues. This method has general applicability for both the production of borrelidin and borrelidin analogues.

On the basis of this finding, the feeding experiments with alternative carboxylic acids were repeated in *S. parvulus* Tü4055/*borG:aac3(IV)*, and extended to include 2,3-dimethyl succinic acid and 2-methylsuccinic acid; the new compounds derived from the incorporation of these alternative starter units, **19** and **20** repectively, are described in figure 10.

In an attempt to improve the titre of borrelidin produced in fermentation cultures of *S. parvulus* Tü4055 through other means, additional copies of the genes *borE* and *borL* were introduced into the organism in vectors that place them under the control of the strong constitutive promoter ermE*. It was anticipated that the over-expression of these genes would increase the intra-cellular levels of the starter acid, which appears to be limiting with respect to borrelidin production.

The genes *borE* and *borL* were amplified by PCR, cloned into the vector pEM4, and then introduced into *S. parvulus* Tü4055 as described in examples 29 and 30 respectively. In addition, the vector pEM4 alone (not containing any insert) was also introduced in *S. parvulus* Tü4055 and used as a control. The resulting strains were

grown, extracted and analysed as described in examples 1 and 4. Introduction of the vector as a control did not significantly effect the levels of borrelidin production. However, the expression of additional copies of either *borE* or *borL* in this manner brought a 4.2±0.3 and 4.3±0.7-fold increase respectively in the titre of borrelidin relative to the wild type strain. Presumably, the steps of biosynthesis catalysed by their gene products are rate limiting, or alternatively their gene products may have a positive regulatory function. For example *borL* shows greatest homology to auxin response proteins from plants. Auxins are hormones involved in the regulation of various cellular processes in plants, and *borL* may represent the first example of a related gene having regulatory function in a bacteria. As controls, an additional copy of *borJ*, *borO* and *borA5*, under the control of ermE* in pEM4, were introduced into *S. parvulus* Tü4055, but did not have any significant effect upon borrelidin titre. This was anticipated as none of the respective gene products are anticipated to be involved in starter unit biosynthesis. In addition, up-regulation of the putative 'stuttering' PKS module (*borA5*) did not increase borrelidin titre, further indicating that iterative use of this module occurs, rather than three independent copies being utilized. The lack of an effect on titre when *borO* is up-regulated indicates that there is most probably no limitation placed upon borrelidin production due to toxicity in the producing organism and so indicates that there is further scope for titre improvement.

Formation of the nitrile moiety at C12

Sequence analysis of the AT domain of the borrelidin PKS module 3 indicates that the substrate utilised for the third round of chain extension is methylmalonyl-CoA. Thus, the carbon atom of the nitrile moiety most probably arises from the methyl group of methylmalonyl-CoA. This was verified by stable isotope feeding experiments. Feeding [2,3-$^{13}C_2$]sodium propionate to *S. parvulus* Tü113 gave borrelidin which displayed intact labelling of the carbons at C4-C24, C6-C25, C8-C26, C10-C27 and C12-C28, and identical specific incorporations (as determined within the limits of our experimental methods), as expected (figure 2). These data indicate that the conversion of the C12-methyl group occurs either during chain assembly at, or after, the incorporation of the third extension unit, or that it occurs after polyketide chain assembly and release from the PKS. Based on functional assignments given to the borrelidin biosynthetic genes, in conjunction with the gene disruption data described in

table 10, both *borI* and *borJ* are clearly implicated in formation of the nitrile moiety at C12, while others such as *borK* may also be.

The cytochrome P450 hydroxylase BorI shares greatest similarity to TylHI, which catalyses the hydroxylation of an exocyclic methyl group of the tylosin macrolactone prior to addition of a deoxyhexose moiety (Fouces *et al.*, 1999). BorI is therefore believed to catalyse oxidation of the C12-methyl group during borrelidin biosynthesis. In agreement with this the *borI⁻* mutant *S. parvulus* Tü4055/*borI::aac3(IV)* fails to produce borrelidin but accumulates a new product **14** (figure 6) that is less polar than borrelidin. **14** is readily transformed to borrelidin when fed to the *borE⁻* mutant *S. parvulus* Tü4055/*borE::aac3(IV)* which lacks the ability to synthesise the PKS starter unit but maintains the rest of the borrelidin biosynthetic genes intact. Fermentation of *S. parvulus* Tü4055/*borI::aac3(IV)* followed by extraction and isolation provided ~30 mg of **14** (example 31). Full structural analysis of **14** identified it as 12-desnitrile-12-methylborrelidin (pre-borrelidin). This is consistent with the proposed role of BorI in borrelidin biosynthesis and provides a route to novel borrelidin analogues with a methyl group attached to C12 of the macrolactone ring.

The putative PLP dependent aminotransferase BorJ is believed to catalyse the introduction of a nitrogen atom into borrelidin at the activated C28-position, probably via a C12-formyl moiety. In agreement with this the *borJ⁻* mutant *S. parvulus* Tü4055/*borJ::aac3(IV)* does not produce borrelidin and accumulates a new compound that is more polar than borrelidin. This new compound is not transformed to borrelidin when fed to mutant *S. parvulus* Tü4055/*borE::aac3(IV)* which indicates that it is probably a shunt metabolite rather than an intermediate in borrelidin biosynthesis. Fermentation of *S. parvulus* Tü4055/*borJ::aac3(IV)* allowed the isolation of 17 mg of the accumulated compound (example 32). Detailed structural analysis identified the accumulant as 12-desnitrile-12-carboxyl borrelidin **2**.

In addition to the compounds isolated from mutation of the borrelidin biosynthetic genes, 12-desnitrile-12-formyl borrelidin **15** is isolated from the fermentation supernatant of *S. parvulus* Tü113. The fermentation media and conditions used for these experiments differ from those we have described so far herein, but are designed to maximise the production of borrelidin. We propose that this altered medium, in combination with a drop in the dissolved oxygen concentration that is observed to occur during this specific fermentation, promoted the accumulation of **15**. **15** is readily transformed to borrelidin when fed to the mutant *S. parvulus*

Tü4055/*borE::aac3(IV)* which lacks the ability to synthesise the PKS starter unit but maintains the rest of the borrelidin biosynthetic genes intact.

The above data lead us to propose a biosynthetic route to the nitrile moiety of borrelidin as presented in figure 6. The C12-methyl carbon of pre-borrelidin **14** is first oxidised by BorI to introduce an allylic hydroxyl group at C28 **(16)**. This hydroxyl group is then converted to the formyl moiety attached to C12 **(15)** using a method selected from the group comprising: spontaneous oxidation (including oxidation mediated by some background enzyme) the action of a specific gene of the borrelidin biosynthetic gene cluster; candidate gene products are thus BorI itself, acting in a multifunctional manner and operating via the formation of a gem-diol structure at C12 followed by dehydration; or alternatively, via one of the oxidoreductase encoding genes such as *borC* or bor*K*. The next step is anticipated to be BorJ-catalysed transamination of **15** in order to introduce a nitrogen atom at C28, in the form of an amine, through a pyridoxamine phosphate mediated process. The putative product amine **17** then undergoes oxidation, possibly spontaneously, but most probably by an enzymic activity such as BorI (certain parallels can be drawn to the biosynthesis of nitriles in plants (Celenza, 2001; Hahn *et al.*, 1999; Nielson and Møller, 1999)) or by the products of one of the oxidoreductase encoding genes, e.g. *borC or borK*, or by a general oxidoreductase within the proteome.

In order to examine this proposed pathway in more detail a number of biotransformation experiments were performed using pre-borrelidin **14** as substrate for investigating the action of *borI-K* individually and in combination, using pEM4 as vector and *S. albus* J1074 (Chater & Wilde, 1980) as an expression strain. Expression of *borI* or *borJ* individually did not give borrelidin production on addition of **14**. The added **14** was only consumed during biotransformation with *borI* (and not in any of the control experiments); the **14** added was identified as being converted to the shunt metabolite **2**. However, co-expression of *borI* & *borJ* did convert the added **14** to borrelidin. It thus appears that either BorI or general proteome activities in *S. albus* are capable of oxidising the proposed amine intermediate **17** in the borrelidin biosynthetic pathway. In addition to the feeding of pre-borrelidin **14**, 12-desnitrile-12-carboxyl borrelidin **2** was also fed to the three strains described above. No conversion of **2** to borrelidin was observed in any of these experiments, reinforcing the idea that **2** is a shunt metabolite.

Detailed investigation of genomic DNA from three borrelidin producing strains, *S. rochei* ATCC23956, *S. parvulus* Tü113 and *S. parvulus* Tü4055, using numerous

restriction digests and subsequent Southern Blot analysis, indicates that the borrelidin biosynthetic gene clusters of these three organisms are very closely conserved. It therefore appears that the borrelidin biosynthetic pathways of these strains are very similar. This assumption allows us to consider the data above, which are obtained from different strains, as applicable to a single biosynthetic pathway.

It is clear to one skilled in the art that manipulation of the genes involved in formation of the C12-nitrile moiety of borrelidin, for example *borI*, or *borJ*, is a generally useful method for the production of novel borrelidin related molecules and borrelidin derivatives with altered functionality at C12. In addition, the transfer of these genes to other organisms producing other natural or engineered polyketide products may allow the incorporation of nitrile moieties into such compounds.

In an extension of this work, disruptions in *borI* and *borJ* are separately made in the strain *S. parvulus* Tü4055/*borG:aac3(IV)* to give the doubly mutated strains *S. parvulus* Tü4055/*borG:aac3(IV)/borI::hyg* and *S. parvulus* Tü4055/*borG:aac3(IV)/borJ::hyg* (examples 27 & 28 respectively). These strains are fed alternative carboxylic acids, *trans*-cyclobutane-1,2-dicarboxylic acid, 2,3-dimethylsuccinic acid and 2-methylsuccinic acid, (as described above) and are found to produce the mutasynthetic borrelidin analogues carrying, either, a methyl (**21**, **22** and **23** respectively) or a carboxyl function at C12 (**24**, **25** and **26** respectively) in place of the nitrile group, and which are also derived from alternative starter units corresponding to the exogenously supplied carboxylic acids. This orthogonal library of new compounds is described in figure 11 and the observed UV chromophores and mass spectral data for each compound is shown.

<u>Other genes involved in borrelidin production</u>

In addition to the type-I terminal thioesterase domain of the borrelidin PKS, a discrete type-II thioesterase is located at the upstream boundary of the biosynthetic gene cluster and is encoded by the gene *borB*. Such discrete type-II TE proteins are commonly found to be associated with type-I PKSs and are believed to play a role in the 'editing' of PKSs by the removal of short chain acyl groups that are formed by unwanted decarboxylation of extender units attached to KS domains (Heathcote *et al.*, 2001). The disruption of such discrete type-II TEs in the picromycin (Xue *et al.*, 1998) and tylosin (Butler *et al.*, 1999) biosynthetic clusters leads to a significant reduction in

titre of both macrolides. In accordance with these results, disruption of *borB* (example 13) gave a mutant that produced between 43-75% of the parental wild type titre.

The self-resistance of *S. parvulus* strains to borrelidin is most probably due to the product of *borO*, which encodes a threonyl tRNA synthetase homologue. Threonyl-tRNA synthetase is the molecular target of borrelidin in sensitive strains (Paetz & Ness, 1973). It is predicted that BorO is resistant to the action of borrelidin, and acts to produce threonyl-tRNAs in cells that make borrelidin, effectively complementing the normal threonyl-tRNA which are inhibited. To verify this hypothesis *borO* was amplified by PCR and cloned in to the expression vector pEM4A, which puts *borO* under the control of the strong constitutive promoter ermE* (example 26). The resulting vector p*borOR* was then transformed into the borrelidin-sensitive strain *Streptomyces albus* J1074 (Chater & Wilde, 1980). Comparison of this strain with that containing only the expression vector pEM4A, using a soaked disk bioassay, clearly indicated that expression of *borO* confers resistance to borrelidin.

EXAMPLES

General methods

Restriction enzymes, other molecular biology reagents, antibiotics and chemicals were purchased from standard commercial sources. Restriction endonuclease digestion and ligation followed standard methods (Sambrook, J. *et al.*, 1989).

Example 1: Fermentation of *S. parvulus* strains

The following method is generally useful for culturing S. parvulus *for the production of borrelidin and/or borrelidin analogues*:

A seed flask containing NYG medium (30 ml in a 250 ml Erlenmeyer flask) was inoculated from a working stock (0.5 ml). NYG medium contains, in deionised water: beef extract (0.3 %), Bacto peptone (0.5 %), glucose (1 %) and yeast extract (0.5 %). After 2 days shaking in a rotary incubator (2-inch throw; 30°C; 250 rpm) the resulting cream culture was used to inoculate PYDG production medium (30 ml in a 250 ml Erlenmyer flask; 10 % innoculum). PYDG medium contains per litre of deionised water: peptonised milk nutrient (1.5 %), yeast autolysate (0.15 %), dextrin (4.5 %) and glucose (0.5 %) adjusted to pH 7.0. After 5 days shaking on a rotary incubator (2-inch throw; 30°C; 250 rpm) the culture was harvested for analysis as described in example 4, or for

isolation purposes as required. For quantitative analysis these experiments were performed in triplicate.

The following method is useful for the feeding of exogenous carboxylic acids to S. parvulus *strains*:

The *S. parvulus* strain was grown as described above. After 24 hours growth in PYDG production medium, the carboxylic acid of choice was added as a 50 µl single aliquot (0.6 M solution in 70 % methanol after neutralization with 5 N NaOH). The resulting culture was harvested after 5 days total fermentation and analysed as described in example 4. For quantitative studies these experiments were performed in triplicate, and the equivalent fed and unfed WT strains served as controls.

<u>Example 2: Cryopreservation of *S. parvulus* strains</u>

Working stocks

Working stocks of vegetative mycelia were prepared by mixing a 2 day old seed culture grown in NGY medium (0.5 ml) with cryopreservative (0.5 ml). Cryopreservative consists of 20 % glycerol and 10 % lactose in deionised water.

Spore stocks

Strains of *S. parvulus* were incubated on HA agar plates at 30°C. After 14 days the resulting spores from a single plate were harvested and suspended in of cryopreservative (1 ml). HA agar contains in deionised water: 0.4% yeast extract, 1% malt extract, 0.4 % dextrose and 1.5 % agar adjusted to pH 7.3.

<u>Example 3: Cloning of the borrelidin biosynthetic gene cluster and disruption of *borA2* & *borA3*</u>

Cosmid library generation

A cosmid library was constructed in pWE15 cosmid vector using the Gigapack® III Gold Packaging Extract kit according to the manufacturer's handbook (Stratagene). Chromosomal DNA was extracted from *S. parvulus* Tü4055 according to standard protocols (Kieser *et al.*, 2000) and treated with *Sau*3AI prior to cloning into pWE15. A number of the resulting *E. coli* transformants (3300) were picked and transferred to 96 well microtitre plates containing Luria Broth (LB) medium (0.1 ml per well) with ampicillin (100 µg/ml). The resulting clones were replica-plated to Luria agar (LA)

plates containing ampicillin (100 µg/ml). After incubation overnight at 37°C colonies were transferred to nylon membrane filters for *in situ* colony hybridization analysis according to published protocols (Sambrook *et al.*, 1989).

Library screening

The cosmid library was screened using a probe that was generated using the DIG DNA Labelling and detection kit (Roche) according to the manufacturers instructions. The probe used was a *Bgl*II-*Bam*HI fragment (1.7 kbp) obtained from the gene that encodes module 6 of the third subunit of the oleandomycin PKS from *Streptomyces antibioticus* (Swan *et al.*, 1994).

Disruption of the borrelidin biosynthetic gene cluster

Cosmids that gave a positive response when screened as described above were digested with *Bam*HI and fragments of less than 3 kbp were subcloned into pOJ260 (Bierman *et al.*, 1992). These were then used to transform protoplasts of *S. parvulus* Tü4055 as described in example 5. The resulting transformants were then assessed for the ability to produce borrelidin. Two clones were borrelidin non-producers; both were obtained from cosBor32A2 and contain sequence typical of a modular PKS. The remaining cosmids were then screened using probes obtained from the two *Bam*HI fragments, which led to the identification of the overlapping cosmid cosBor19B9 that contained the remainder of the borrelidin biosynthetic cluster.

Sequencing of cosBor32A2 and cosBor19B9

The cosmids cosBor32A2 and cosBor19B9 were transformed into *E. coli* DH10B and the resulting clones grown at 37°C in 2xTY media (30 ml) containing ampicillin. After 15 hours the cells were harvested and Qiagen Tip 100 kits were used to prepare cosmid DNA. Approximately 5 µg of the cosmid DNA was digested with *Sau3*AI (1 U). Samples were taken at 2, 4, 6, 8 & 10 minute intervals after the enzyme was added and quenched into an equal volume of ice cold 0.5M EDTA. The samples were mixed and then analysed by gel electrophoresis, and those fragments between 1.5-2.0 kbp recovered from the gel. The fragments were cloned into linearised and dephosphorylated pHSG397 (Takeshita *et al.*, 1987), and transformed into *E. coli* DH10B. The resulting clones that contained insert were grown in 2xTY medium (2 ml) containing chloramphenicol (30 µg/ml) and purified using Wizard kits (Promega).

DNA sequencing was carried out using an Applied Biosystems 800 Molecular Biology CATALYST robot to perform the dideoxy terminator reactions, which were then loaded into an ABI Prism 3700 automated sequencer (Applied Biosystems). The raw sequence data was processed using the Staden software package. Assembly and contig editing was performed using GAP (Genome Assembly Program) version 4.2 (Bonfield *et al.*, 1995). The GCG package (Devereux *et al.*, 1984) version 10.0 was used for sequence analysis.

Example 4: Chemical analysis of *S. parvulus* strains

The following method is useful for analysing fermentations (see example 1) for the production of natural borrelidins and of engineered borrelidin analogues:

In a 2 ml Eppendorf tube, an aliquot of 5 day old fermentation broth (1 ml) was adjusted to pH~3 by the addition of 90 % formic acid (ca. 20 µl). Ethyl acetate (1 ml) was added to the sample and mixed vigorously for 10 min using a vortex tray. The mixture was separated by centrifugation in a microfuge and the upper phase removed to a clean 2 ml Eppendorf tube. The ethyl acetate was removed by evaporation using a Speed-Vac. Residues were dissolved into methanol (250 µl) and clarified using a microfuge. Analysis was performed on an Agilent HP1100 HPLC system as described below:

Injection volume:	50 µl
Column stationary phase:	150 x 4.6 mm column, base-deactivated reversed phase silica gel, 3 µm particle size (Hypersil C_{18}-BDS).
Mobile phase A:	10 % acetonitrile:90 % water, containing 10 mM ammonium acetate and 0.1 % TFA.
Mobile phase B:	90 % acetonitrile:10 % water, containing 10 mM ammonium acetate and 0.1 % TFA.
Mobile phase gradient:	T=0 min, 25%B; T=15, 100%B; T=19, 100%B; T=19.5, 25%B; T=25, 25%B.
Flow rate:	1 ml/min.
Detection:	UV at 258 nm (DAD acquisition over 190-600 nm); MS detection by electrospray ionisation over *m/z* range 100-1000 amu, with +/-ve ion mode switching.

Example 5: Protoplast transformation protocol for *S. parvulus* Tü4055

A seed flask containing tryptone soy broth (TSB) medium (10 ml in a 100 ml Erlenmyer flask) was inoculated from a working stock (0.15 ml). After 3 days shaking on a rotary incubator (30°C, 250 rpm), 5 ml of the culture was used to inoculate R5 medium (Kieser *et al.*, 2000) (50 ml in a 250 ml Erlenmeyer flask) that was then shaken on a rotary incubator for 24 hours (30°C, 250 rpm). The PEG mediated transformation of protoplasts was then performed according to standard published protocols (Kieser *et al.*, 2000).

Example 6: Replacement of borAT4 with rapAT2 - production of C10-desmethyl borrelidin

The borrelidin PKS AT4 domain is replaced with the AT2 domain of the rapamycin polyketide synthase as follows:

CosBor32A2 is digested with *Eco*RI and the 5429 bp band isolated. This is used as a template for PCR using the oligos CM410 (5'-AAAATGCATTCGGCCTGAACGGCCCCGCTGTCA-3') (SEQ ID No.44) and CM411 (5'-AAATGGCCAGCGAACACCAACACCACACCACCA-3') (SEQ ID No.45). CM410 introduces an *Nsi*I restriction site for cloning purposes and CM411 introduces an *Msc*I site for use in the introduction of a heterologous AT. The ~1.1 kbp product is cloned into pUC18 digested with *Sma*I and dephosphorylated. The insert can ligate in two orientations and the reverse orientation is screened for by restriction enzyme analysis and the insert sequenced. One correct plasmid is designated pCJM462. Methylation deficient DNA (specifically dcm$^-$) of pCJM462 and pCJR26 (Rowe *et al.* 1998) is isolated by passaging the plasmids through *E. coli* ET12567. Each plasmid is then digested with *Msc*I and *Xba*I and the ~7.8 kbp fragment from pCJR26, containing the rapamycin AT2 and sequences downstream in pCJR26, is ligated to the ~3.8 kbp backbone generated by digestion of pCJM462. Plasmid pCJM463 is identified by restriction analysis.

CosBor32A2 is digested with *Eco*RI and *Eco*RV and the 2871 bp band isolated. This is used as a template for PCR using the oligos CM412 (5'-AAAGTCCTAGGCGGCGGCCGGCGGGTCGACCT-3') (SEQ ID No.46) and CM413 (5'-TTTAGATCTCGCGACGTCGCACGCGCCGAACGTCA-3') (SEQ ID No.47). CM412 introduces an *Avr*II restriction site that joins, in frame, the downstream borrelidin

homology to the heterologous AT, and CM413 introduces a *Bgl*II site for cloning purposes. The ~1.1 kbp product is cloned into pUC18 digested with *Sma*I and dephosphorylated. The insert can ligate in two orientations and the reverse orientation is screened for by restriction enzyme analysis and the insert sequenced. One correct plasmid is designated pCJM464.

Plasmids pCJM463 and pCJM464 are digested with *Avr*II and *Xba*I and the ~1.1 kbp fragment from pCJM464 is ligated into the ~4.7 kbp backbone of pCJM463 to give pCJM465, which is identified by restriction enzyme analysis. pCJM465 contains the hybrid rapamycin AT2 with flanking regions of borrelidin sequence which provide homology for integration and secondary recombination.

Plasmid pCJM465 is digested with *Nsi*I and *Bgl*II and the ~3 kbp fragment is cloned into pSL1180 previously digested with *Nsi*I and *Bam*HI to give pCJM466. Plasmid pCJM466 is then digested with *Nsi*I and the apramycin cassette is incorporated on a *Pst*I fragment from pEFBA (Lozano *et al.* 2000) to give the replacement vector pCJM467. pCJM467 is introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin (25 μg/ml) are initially identified, and then passaged several times through MA media without antibiotic selection in order to promote the second recombination (Fernandez *et al.* 1998). Several apramycin-sensitive colonies are isolated and analysed by PCR and Southern blot. The new mutant is named *S. parvulus* Tü4055/467.

S. parvulus Tü4055/467 is analysed as described in example 1 and shown to produce a mixture compounds with the correct UV spectrum. One of the new major components that is more polar than borrelidin has the correct retention time for 10-desmethyl borrelidin **3**. LCMS analysis indicates an *m/z* ratio for a compound that is 14 mass units lower than borrelidin as expected, and with an appropriate mass fragmentation pattern. Borrelidin itself is also produced, but at levels lower than the WT organism.

<u>Example 7: Mutation of the methylmalonyl-CoA selective motif of borAT4 to generate 10-desmethyl borrelidin</u>

Site directed mutagenesis of acyl transferase domains may also be used to alter the specificity of an AT. In this example the specificity of borAT4 is directed from methyl-malonyl-CoA towards malonyl-CoA. An amino acid motif has been identified (Reeves *et al.*, 2001; WO 02/14482) which directs the specificity of an AT. The motif

YASH, as observed in borAT4, is found in methylmalonyl-CoA specific ATs and in this example it is altered to HAFH which is found in malonyl-CoA specific ATs.

CosBor32A2 is digested with *Nco*I and the 5167 bp band isolated. This is used as a template for PCR using the primers CM414 (5'-AAACTGCAGAGTCGAACATCGGTCACACGCAGGC-3') (SEQ ID No.48) and CM415 (5'-AAAATGCATGATCCACATCGATACGACGCGCCCGA-3') (SEQ ID No.49). CM414 introduces a *Pst*I restriction site for cloning purposes, and CM415 is a mutagenic primer covering the motif encoding region of the AT which will effect the amino acid changes and contains an *Nsi*I site for cloning purposes. The ~1.1 kbp fragment is cloned into pUC18 digested with *Sma*I and dephosphorylated. The insert can ligate in either orientation and the forward orientation is screened for by restriction enzyme analysis and the insert sequenced. One correct plasmid is designated pCJM468.

A second PCR reaction is performed using the 5167 bp *Nco*I fragment of CosBor32A2 and the primers CM416 (5'-TAAATGCATTCCATTCGGTGCAGGTGGAGTTGATCC-3') (SEQ ID No.50) and CM417 (5'-ATAGGATCCCCTCCGGGTGCTCCAGACCGGCCACCC-3') (SEQ ID No.51). CM416 introduces an *Nsi*I restriction site and is also a mutagenic primer covering the motif encoding region of the AT, and CM417 introduces a *Bam*HI site for cloning purposes. The ~1.1 kbp fragment is cloned into pUC18 previously digested with *Sma*I and dephosphorylated. The insert can ligate in two orientations and the forward orientation is screened for by restriction enzyme analysis and the insert sequenced. One correct plasmid is designated pCJM469.

Plasmids pCJM468 and pCJM469 are digested with *Nsi*I and *Xba*I and the ~1.1 kbp fragment from pCJM468 is ligated into the ~3.8 kbp backbone of pCJM469 to give pCJM470, which is identified by restriction enzyme analysis. pCJM470 contains the mutated motif of borAT4 with ~1.1 kbp of homologous DNA on either side which provide homology for integration and secondary recombination.

Plasmid pCJM470 is digested with *Pst*I and *Bam*HI and the ~2.2 kbp fragment is cloned into pSL1180 (Amersham Biosciences) previously digested with *Pst*I and *Bam*HI to give pCJM471. Plasmid pCJM471 is then digested with *Pst*I and the apramycin cassette is incorporated on a *Pst*I fragment from pEFBA (Lozano *et al.*, 2000) to provide the replacement vector pCJM472.

The replacement vector pCJM472 is introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin

are initially identified, and then passaged several times through MA media without antibiotic selection in order to promote the second recombination (Fernandez *et al.*, 1998). Several apramycin-sensitive colonies are isolated and analysed by PCR and Southern blot, and one is selected that contains the new AT4 sequence containing the mutated motif and the *Nsi*I site. The new mutant is named *S. parvulus* Tü4055/472.

S. parvulus Tü4055/472 is grown and analysed as described in example 1 and shown to produce a mixture of compounds with the correct UV profile for borrelidin. One of the new major components, that is more polar than borrelidin, has the correct retention time for authentic **3**. LCMS analysis indicates an *m/z* ratio for a compound that is 14 mass units lower than borrelidin as expected, and with an appropriate mass fragmentation pattern. Borrelidin itself is also produced, but at levels lower than the WT organism.

<u>Example 8: Introduction of the borrelidin loading module into the erythromycin PKS</u>

The borrelidin loading module was amplified for each of the four putative start codons. The PCR template was a 3376 bp *Bam*HI fragment of cosBor32A2 covering the region from nucleotides 15858 to 19234 of SEQ ID No.1. The reverse primer CM368 (5'-TTTCCTGCAGGCCATCCCCACGATCGCGATCGGCT-3') (SEQ ID No:52) introduces a *Sbf*I site at the sequence corresponding to the start of KS1 of *borA2* (conserved MACRL motif) and is used with each of the forward primers CM369 (5'-TTTCATATGACAGGCAGTGCTGTTTCGGCCCCATT-3') (SEQ ID No.53), CM370 (5'-TTTCATATGGCGGATGCCGTACGTGCCGCCGGCGCT-3') (SEQ ID No.54), CM371 (5'-TTTCATATGCCCCAGGCGATCGTCCGCACCAC-3') (SEQ ID No.55) and CM372 (5'-TTTCATATGGTCTCGGCCCCCCACACAAGAGCCCTCCGGGC-3') (SEQ ID No:56). The four PCR products (of 2834, 2720, 2411 and 2117 bp respectively) were cloned into pUC18 that had previously been digested with *Sma*I and dephosphorylated. The resulting plasmids were designated pCJM370, which contains the largest insert, pCJM371, pCJM372 and pCJM373, which contains the smallest insert.

The four borrelidin loading module fragments were introduced into the vector pKS1W, which contains a *Pst*I site at the start of eryKS1 of DEBS1-TE in the conserved MACRL motif (Rowe *et al.*, 2001); *Pst*I gives the same overhang as *Sbf*I. pKS1W is a pT7-based plasmid containing DEBS1-TE on an *Nde*I/*Xba*I fragment, with unique sites flanking the loading module, a unique *Pst*I site at nucleotide position 1698

of the DEBS1-TE encoding gene and a unique *Nde*I site at the start codon. The borrelidin loading module fragments were excised as follows: pCJM370 was digested with *Nde*I and *Sbf*I, pCJM371 and pCJM373 were digested with *Nde*I and *Pst*I, and pCJM372 was digested with *Nde*I, *Pst*I and *Dra*I. Each loading module containing fragment was cloned into pKS1W previously digested with *Nde*I and *Pst*I. The resulting plasmids were designated pCJM384, which contains the largest insert, then pCJM386, pCJM388 and pCJM390, which contains the smallest insert.

The hybrid PKS fragments were transferred into pCJR24, which is a suitable vector for transformation of *S. erythraea* WT and *S. erythraea* DM, and for expression of the resulting hybrid PKS (WO 98/01546). Each loading module construct was excised along with a 2346 bp fragment of DNA from DEBS1 in order to allow integration into the chromosome. In order to achieve this, pCJR24 is digested with *Xba*I and end-filled using Klenow fragment of DNA polymerase I. This is then digested with *Nde*I to give the backbone fragment. Into this, the four hybrid PKS fragments containing the borrelidin loading modules plus the region of DEBS1 sequence for integration are cloned as *Nde*I/*Eco*RV fragments from pCJM384, pCJM386, pCJM388 and pCJM390 to give pCJM400, pCJM401, pCJM402 and pCJM403 respectively.

Plasmids pCJM400, pCJM401, pCJM402 and pCJM403 were introduced into *S. erythraea* by transformation of *S. erythraea* DM protoplasts as described elsewhere (Gaisser *et al.*, 2000). The resulting mutants were analysed by PCR and Southern blot to confirm the presence of the plasmid on the chromosome and to establish that correct integration had occurred. A number of mutants that appeared correct by these methods were grown, extracted and analysed according to standard methods for polyketide production from *S. erythraea* strains (Wilkinson *et al.*, 2000). When compared to control strains using LCMS methods, the extracts from several of these mutants contained new compounds at reasonable levels. Analysis of their MS spectra showed the presence of a compound with m/z = 485.3 ($[M-H]^-$, **6**) in negative ion mode. This is in agreement with the expected product compound (M = 486.3).

<u>Example 9: Fusion of PKS modules 4 and 5 (*S. parvulus* Tü4055/*borA4-A5*)</u>

To examine the iterative action of module 5, the two separate proteins encoding modules 4 and 5 were fused together through manipulation at the genetic level. The fusion was performed by a gene replacement in which the last ~1 kbp of *borA4* and the

first ~1 kbp of *borA5*, were fused by converting the overlapping stop and start codons respectively into an arginine residue, introducing a new *Xba*I site and converting the two separate orfs into one.

In the first step of the mutagenesis, two separate PCR amplifications were performed. In the first PCR reaction, the template DNA was cosBor19B9, and the primers were B1819A (5'-GTCATGCATGCGGCGGGCTC-3') (SEQ ID No.57) and B1819B (5'-GGTCTAGAACGGCCGAACTT-3') (SEQ ID No.58). The 1063 bp product was purified, digested *Nsi*I-*Xba*I and cloned into pSL1180 (Amersham Biosciences) digested similarly to give plasmid pSL18-19AB. The second PCR reaction amplified the *borA5* fragment and used the primers B1819C (5'-GTTCTAGAACCTCGGTCGGC-3') (SEQ ID No.59) and B1819D (5'-CTGGATCCCACGCTGCTGCG-3') (SEQ ID No.60). The 1033 bp product was purified, digested with *Xba*I-*Bam*HI and cloned into pSL18-19AB that had been digested similarly, to give plasmid pSL18-ABCD. Finally, the apramycin cassette from pEFBA (Lozano *et al.*, 2000) was excised as a *Pst*I fragment and cloned into pSL18-19ABCD digested with *Nsi*I to give the replacement vector pSL18-19Apra.

The replacement vector pSL18-19Apra was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin (25 µg/ml) were initially selected, and then passaged several times through MA media without selection. Several apramycin-sensitive colonies were obtained, two of which produced borrelidin while the others did not.

Chromosomal DNA was extracted from all of the apramycin sensitive colonies and checked initially by PCR using the primers BLDA (5'-GGAGACTTACGGGGGATGC-3') (SEQ ID No.61) and BLDB (5'-CTCCAGCAGCGACCAGAAC-3') (SEQ ID No.62) that are selective for the loading module (*borA1*). A 2.9 kbp fragment was observed for the control and the two borrelidin-producing mutants, but not for the non-producing strains. This result is symptomatic and characteristic of non-specific deletions in the chromosome.

The two borrelidin-producing colonies were analysed further by PCR using the primers B19A (5'-CCCATGCATCACCGACATAC-3') (SEQ ID No.63) and B19B (5'-GCGATATCCCGAAGAACGCG-3') (SEQ ID No.64) in order to check the fusion site. The method was as described above. Both the colonies and the controls gave a PCR product of 1010 bp, but upon digestion with *Xba*I only those that carried the fusion-producing mutation gave digestion to 600 and 400 bp fragments. Only one of the

borrelidin-producing colonies harboured the fusion, while the other had reverted to wild type. Final confirmation came from Southern analysis using a *Bam*HI-*Xho*I internal fragment from *borA5* as probe over chromosomal DNA digested with *Xba*I and *Bcl*I. The control and wild type revertant colony showed a fragment of 11.5 kbp as expected, while the fusion mutant showed a fragment of 7.8 kbp as expected. This new mutant was named *S. parvulus* Tü4055/*borA4-A5*. *S. parvulus* Tü4055/*borA4-A5* was shown to produce borrelidin at 26±5% of the WT titre, following the protocol described in example 1.

Example 10: Fusion of PKS modules 5 and 6 (*S. parvulus* Tü4055/*borA5-A6*)

This experiment was performed for the same reason as, and in an analogous manner to, that of example 9 above. The fusion of these orfs introduced an additional leucine residue into the new protein at the fusion point, in addition to a new *Spe*I site at the genetic level. In the first step of the process two PCR fragments were generated using cosBor19B9 as template. The first PCR reaction amplified the *borA5* region and used the primers B1920A (5'-GCCAAGCTTCCTCGACGCGC-3') (SEQ ID No.65) and B1920B (5'-CACTAGTGCCTCACCCAGTT-3') (SEQ ID No.66). The 804 bp product was purified and digested with *Hind*III-*Spe*I. The second PCR reaction amplified the *borA6* region and used the primers B1920C (5'-CACTAGTGACGGCCGAAGCG-3') (SEQ ID No.67) and B1920D (5'-TCGGATCCGTCAGACCGTTC-3') (SEQ ID No.68). The 960 bp product was purified and digested with *Spe*I-*Bam*HI. The two purified and digested gene products were then cloned together into pOJ260 that had been digested with *Hind*III-*Bam*HI to give the replacement vector pOJF19-20. pOJF19-20 was introduced into *S. parvulus* Tü4055 by protoplast transformation to give apramycin resistant colonies. One such colony was passaged several times through MA media without selection in order to promote double recombination. Two apramycin sensitive colonies were obtained, and chromosomal DNA from these was examined by Southern hybridisation to check for the presence of a 3.2 kbp *Bam*HI fragment (to control for unwanted deletions in the loading module) and a 3.4 kbp *Spe*I-*Bam*HI fragment to verify correct introduction of the *borA5-A6* fusion (5.8 kbp *Bam*HI fragment in the WT). One of the apramycin colonies carried the correct mutation without deletion and was named *S. parvulus* Tü4055/*borA5-A6*. *S. parvulus* Tü4055/*borA5-A6* was shown to produce borrelidin at 25±4% of the WT titre, following the protocol as described in example 1.

Example 11: Fusion of PKS modules 4, 5 and 6 (*S. parvulus* Tü4055/*borA4-A5-A6*)

To generate the strain *S. parvulus* Tü4055/*borA4-A5-A6* we took advantage of the previously obtained strain *S. parvulus* Tü4055/*borA4-A5* (Example 9) and plasmid pOJF19-20 (Example 10). pOJF19-20 was introduced into *S. parvulus* Tü4055/*borA4-A5* by protoplast transformation to give apramycin resistant colonies. One such colony was passaged several times through MA media without selection in order to promote double recombination. One apramycin sensitive colony was obtained, and chromosomal DNA from it was examined by Southern hybridisation to check for the presence of a 3.2 kbp *BamHI* fragment (to control for unwanted deletions in the loading module), a 3.4 kbp *SpeI-BamHI* fragment to verify correct introduction of the *borA5-A6* fusion (5.8 kbp *BamHI* fragment in the WT) and a 6.4 kbp *SpeI-XbaI* to verify the presence of both fusions, *borA4-A5* and *borA5-A6*, within the same strain. The chosen colony carried the correct mutation without deletion and was named *S. parvulus* Tü4055/*borA4-A5-A6*. *S. parvulus* Tü4055/*borA4-A5-A6* was shown to produce borrelidin at 18±5% of the WT titre, following the protocol as described in example 1.

Example 12: Replacement of the erythromycin PKS module 4 with module 5 of the borrelidin PKS – production of ring expanded macrolides

Example 12 describes the replacement of erythromycin module 4 with borrelidin module 5. Borrelidin module 5 is believed to be responsible for three rounds of condensation of methylmalonyl-CoA, in an iterative fashion, within the borrelidin PKS. Previously, erythromycin module 4 has been shown to occasionally act in an iterative fashion 'mis'-incorporating a second methylmalonyl-CoA to make very small amounts of a 16-membered macrolide from the erythromycin PKS. A strain in which the erythromycin module 4 is replaced by borrelidin module 5 is engineered by a replacement strategy as follows, and is based on a derivative process as described for module insertion into the erythromycin PKS (Rowe *et al.*, 2001):

Initially a series of plasmids are made in order to generate a plasmid in which the borrelidin module 5 is flanked by appropriate regions of homology from the erythromycin PKS. In order to facilitate this, the *SbfI* site is first removed from the polylinker of pUC18 by digestion with *PstI*, end-polishing with T4 polymerase and religation. The new plasmid, pCJM409 is identified by restriction enzyme digestion.

Borrelidin module 5 is isolated on an *Sbf*I fragment by ligating together 4 PCR fragments. PCRA is generated by amplification of ~1.4 kb of the beginning of borrelidin module 5 using the 6062 bp *Xcm*I fragment of cosBor19B9 as the template and primers CM384 (5'-AACCTGCAGGTACCCCGGTGGGTGCGGTCGCCCGA-3') (SEQ ID No.69) and CM385 (5'-CGCCGCACGCGTCGAAGCCAACGA-3') (SEQ ID No.70). CM384 introduces an *Sbf*I site in the conserved amino acid sequence MxCR at the beginning of borrelidin module 5. CM385 incorporates a naturally occurring *Mlu*I site that is used in the cloning strategy. PCRA is treated with T4 polynucleotide kinase (T4 PNK, NEB) and cloned into pCJM409 previously digested with *Sma*I and dephosphorylated with Shrimp Alkaline Phosphatase (SAP, Roche). Inserts cloned in the forward direction are screened for by restriction enzyme digestion, and for one correct clone the insert is verified by sequencing. This plasmid is designated pCJM410.

PCRB is generated by amplification of the adjacent ~1.4 kb of borrelidin module 5 using the 6062 bp *Xcm*I fragment of cosBor19B9 as the template and primers CM386 (5'-TGTGGGCTGGTCGTTGGCTTCGAC-3') (SEQ ID No.71) and CM387 (5'-GGTGCCTGCAGCGTGAGTTCCTCGACGGATCCGA-3') (SEQ ID No.72). CM386 binds upstream of the same *Mlu*I site as CM385 contains, which is used in the cloning strategy. CM387 is used to remove the *Sbf*I site within the borrelidin PKS module 5 whilst leaving the overlapping *Pst*I site for cloning. PCRB is treated with T4 PNK and cloned into pCJM409 previously digested with *Sma*I and dephosphorylated with SAP. Inserts cloned in the forward direction are screened for by restriction enzyme digestion, and for one correct clone the insert is verified by sequencing. This plasmid is designated pCJM411.

PCRC is generated by amplification of the downstream adjacent ~1.5 kb of borrelidin module 5 using the 6062 bp *Xcm*I fragment of cosBor19B9 as the template and oligonucleotides CM388 (5'-GAGGAACTCACCCTGCAGGCACCGCT-3') (SEQ ID No.73) and CM395 (5'-CGAACGTCCAGCCCTCGGGCATGCGT-3') (SEQ ID No.74). CM388 binds at the same *Sbf*I site as CM387, but is not mutagenic and retains the *Sbf*I site. CM395 incorporates an *Sph*I site for cloning purposes. PCRC is treated with T4 PNK and cloned into pCJM409 previously digested with *Sma*I and dephosphorylated with SAP. Inserts cloned in the forward direction are screened for by restriction enzyme digestion and for one correct clone the insert is verified by sequencing. This plasmid is designated pCJM412.

PCRD is generated by amplification of the downstream adjacent ~2.1 kb of borrelidin module 5 using the 7211 bp *Bbv*CI fragment of cosBor19B9 as the template and primers CM396 (5'-TGGCACGCATGCCCGAGGGCTGGACGTT-3') (SEQ ID No.75) and CM397 (5'-TTTCCTGCAGGCCATGCCGACGATCGCGACAGGCT-3') (SEQ ID No.76). CM396 contains the *Sph*I site for cloning purposes, and CM397 introduces an *Sbf*I site in the conserved amino acid sequence MxCR at the end of borrelidin module 5. PCRD is treated with T4 PNK and cloned into pCJM409 previously digested with *Sma*I and dephosphorylated with SAP. Inserts cloned in the forward direction are screened for by restriction enzyme digestion, and for one correct clone the insert is verified by sequencing, this plasmid is designated pCJM413.

The four PCR products (PCRA-D) are used to construct the borrelidin module 5 on an *Sbf*I fragment as follows:

pCJM412 is digested with *Sph*I and the ~1.5 kb fragment isolated is cloned into pCJM413 previously digested with *Sph*I and dephosphorylated with SAP. This gives plasmid pCJM414, which is identified by restriction enzyme digestion.

pCJM414 is digested with *Sbf*I and the ~3.6 kb fragment isolated is cloned into pCJM411 previously digested with *Pst*I and dephosphorylated with SAP. This gives pCJM415 which is identified by restriction enzyme digestion.

pCJM410 is digested with *Mlu*I and *Hind*III and the ~1.4 kb fragment isolated is cloned into pCJM415 previously digested with *Mlu*I and *Hind*III. This gives pCJM416, which is identified by restriction enzyme digestion. pCJM416 is a pUC18-based plasmid containing the borrelidin module 5 as an *Sbf*I fragment.

In order to introduce the Borrelidin module 5 into the erythromycin PKS by a replacement strategy, flanking regions of homology from the erythromycin PKS are incorporated for recombination as follows:

PCRE is generated by amplification of ~3.3 kb of the erythromycin PKS directly upstream of the module 4 KS using the 6428 bp *Xmn*I fragment of pIB023 as the template and primers CM398 (5'-AAACATATGGTCCTGGCGCTGCGCAACGGGGAACTG-3') (SEQ ID No.77) and CM399 (5'-TTTCCTGCAGGCGATGCCGACGATGGCGATGGGCT-3') (SEQ ID No.78). CM398 contains an *Nde*I site for cloning purposes and CM399 introduces an *Sbf*I site in the conserved amino acid sequence M/IxCR at the beginning of erythromycin module 4. PCRE is treated with T4 PNK and cloned into pCJM409 previously digested with *Sma*I and dephosphorylated with SAP. Inserts cloned in the

forward direction are screened for by restriction enzyme digestion, and for one correct clone the insert is verified by sequencing, this plasmid is designated pCJM417.

PCRF is generated by amplification of ~3.4 kb of the erythromycin PKS directly downstream of the module 5 KS using the 7875 bp *XmnI/NheI* fragment of pIB023 as the template and primers CM400 (5'-AAACCTGCAGGTTCCCCGGCGACGTGGACTCGCCGGAGTCGTT-3') (SEQ ID No.79) and CM401 (5'-TTTTCTAGAGCGACGTCGCAGGCGGCGATGGTCACGCCCGT-3') (SEQ ID No.80). CM400 introduces an *SbfI* site in the conserved amino acid sequence M/IxCR at the beginning of erythromycin module 4, and primer CM401contains an *XbaI* site for cloning purposes. PCRF is treated with T4 PNK and cloned into pCJM409 previously digested with *SmaI* and dephosphorylated with SAP. Inserts cloned in the forward direction are screened for by restriction enzyme digestion, and for one correct clone the insert is verified by sequencing. This plasmid is designated pCJM418.

pCJM417 is digested with *NdeI* and *SbfI* and the ~3.3 kb fragment is cloned into pCJM418 digested with *NdeI* and *SbfI* (~5.8 kbp) to give pCJM419 which is identified by its restriction digest pattern. pCJM419 contains a unique *SbfI* site which can be used to accept any complete module with *SbfI* (or *PstI*) flanking sites appropriate to place, in-frame, the in-coming module exactly into the conserved region of the KS domain.

The borrelidin module 5 with flanking *SbfI* sites is cloned from pCJM416 as an *SbfI* fragment into the unique *SbfI* site of pCJM419 (which has been dephosphorylated with SAP) to give pCJM420, which is identified by restriction enzyme analysis to confirm the presence and correct orientation of the insert. pCJM420 thus contains borrelidin module 5 with flanking regions of homology to introduce it in-frame between modules 3 and 5 of the erythromycin PKS. The complete insert is removed as an *NdeI/XbaI* fragment from pCJM420 and cloned into pCJM24 digested with *NdeI* and *XbaI* to give the final plasmid pCJM421. pCJR24, and consequently pCJM421, contain an appropriate resistance marker for selection of *S. erythraea* transformants.

Plasmid pCJM421 is used to transform *S. erythraea* strains NRRL2338 (wild type), and *S. erythraea* DM (*eryCIII*⁻, *eryBV*⁻) protoplasts (Yamamoto *et al.*, 1986; Rowe *et al.*, 1998). Integrants are selected for resistance to thiostrepton (50mg/L) and a number of integrants (typically 5-8) are analysed further by Southern blot to confirm that the strains are correct and to identify the site of integration. Two correct integrants

in each case are sub-cultured in TSB liquid media without antibiotic selection in order to promote the second recombination. Several thiostrepton-sensitive colonies are isolated and analysed by PCR and Southern blot, and in each case one selected that contains the new module correctly inserted. This leads to strains *S. erythraea* WT/421 and *S. erythraea* DM/421.

Strain *S. erythraea* DM/421 is cultured under conditions appropriate for the production of erythronolides (Wilkinson *et al.*, 2000). Analysis of fermentation broth extracts using LCMS methods indicates the presence of two new significant peaks when compared to the control strain, and which are less polar than erythronolide B. These have an *m/z* of 435.5 (MNa^+) and 477.5 (MNa^+) respectively, which is consistent with the production of new ring expanded erythronolide B analogues. The compound with *m/z* = 435.5 (**7**) is consistent with the presence of the 16-membered ring-expanded erythronolide B related macrolide reported previously as a minor component of *S. erythraea* WT fermentations (Wilkinson *et al.*, 2000); the compound with *m/z* = 477.5 (**8**) is consistent with the presence of an 18-membered, doubly ring-expanded erythronolide B related macrolide (see figure 8). It is clear to one skilled in the art that such new products can be converted to antibacterial molecules by biotransformation with an appropriate organism, or through the fermentation of the strain *S. erythraea* WT/421. It is further clear to one skilled in the art that the inclusion of such a module into other positions of the erythromycin PKS or into other PKSs may allow the production of novel, ring expanded polyketides in a similar manner.

An alternative strategy for generating this hybrid PKS is to incorporate the borrelidin module 5 in place of erythromycin module 4 within a large plasmid that contains the entire hybrid PKS, followed by transformation of an *eryA*⁻ *S. erythraea* strain. Such an appropriate existing *eryA*⁻ is *S. erythraea* JC2 (Rowe *et al.*, 1998) and the plasmid containing the *eryA* genes under the *actI* promoter, pIB023 that also contains a thiostrepton resistance gene and the *actII*-ORF4 activator. This strategy is accomplished as follows:

pIB023 is digested with *Nde*I and *Bsm*I and the 13.4 kbp fragment is cloned into pCJM419 digested with *Nde*I and *Bsm*I to give plasmid pCJM425. pIB023 is digested with *Bbv*CI and *Xba*I and the approx. 6 kbp fragment is cloned into pCJM425 digested with *Bbv*CI and *Xba*I to give plasmid pCJM426. The *NdeI/Xba*I fragment from pCJM426 is cloned into pCJM395 digested with *Nde*I and *Xba*I. pCJM395 is a plasmid made by

digesting pCJR24 with *SbfI*, end-polishing with T4 polymerase and religating, to give a version of pCJR24 that does not cut with *SbfI*. The resulting plasmid, pCJM427, contains an engineered version of the erythromycin PKS in which module 4 is removed. This backbone is then ready to accept any complete module with appropriate flanking sites (*SbfI* or *PstI*) to generate a hybrid PKS. Introduction of the single borrelidin module 5 is accomplished by digesting pCJM427 with *SbfI*, dephosphorylating the backbone with SAP, and ligating in the *SbfI* fragment from pCJM416, to give pCJM430.

Plasmid pCJM430 is used to transform *S. erythraea* JC2. Integrants are selected for resistance to thiostrepton (50mg/L) and a number of integrants (typically 5-8) are analysed further by Southern blot to confirm that the strains are correct and to identify the site of integration. The resulting correct strain *S. erythraea* JC2/430 is cultured under conditions appropriate for the production of erythromycins (Wilkinson *et al.*, 2000) and analysed for the production of novel compounds **7** & **8**.

Example 13: Disruption of *borB* (*S. parvulus* Tü4055/*borB::aac3(IV)*)

In order to disrupt *borB*, an region of 2751 bp containing *borB* was amplified by PCR using primers B5B (5-'AACTAGTCCGCAGTGGACCG-3') (SEQ ID No.91) and B5A (5'-TCGATATCCTCACCGCCCGT-3') (SEQ ID No.92) and cosmid Bor32A2 as template. The PCR product was purified and then digested at the flanking sites *Spe*I-*Eco*RV and subcloned into pSL1180 digested with the same restriction enzymes to generate pSLB. A *Spe*I-*Age*I fragment (the latter site internal to the insert) from pSLB containing the 5'-end of *borB* was subcloned into the *Spe*I-*XmaI* sites of pEFBA, upstream of the apramycin resistance gene *aac(3)IV*, to produce pEB1. A *Bsa*AI-*Eco*RV fragment (the former site internal to the insert) from pSLB containing the 3'-end of *borB* was then subcloned in the correct orientation into the *Eco*RV site of pEB1 downstream of *aac(3)IV*, to generate pEB2. In this way a 741 bp *Age*I-*Bsa*AI fragment internal to *borB* was deleted and replaced by *aac(3)IV*. Finally, the *Spe*I-*Eco*RV fragment was rescued from pEB2 and subcloned, together with a *Pst*I-*Spe*I fragment containing the *hyg* gene from pLHyg, into the *Pst*I-*Eco*RV sites of pSL1180 to generate pSLBr1. This approach was used in order to avoid possible polar effects.

The vector pSLBr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected, and then passaged several times through MA media without selection. The

replacement was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borB::aac3(IV)*. Strain *S. parvulus* Tü4055/*borB::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was observed and compared to a wild type control. In addition *S. parvulus* Tü4055/*borB::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

Example 14: Disruption of *borC* (*S. parvulus* Tü4055/*borC::aac3(IV)*)

In order to disrupt *borC*, an region of 3553 bp containing *borC* was amplified by PCR using primers B6B (5'-AACTAGTGTGGCAGACGGTC-3') (SEQ ID No.93) and B5A (5'-TCGATATCCTCACCGCCCGT-3') (SEQ ID No.94) and cosmid Bor32A2 as template. The PCR product was purified and then digested with *Spe*I-*Eco*RV and subcloned into the same restriction sites of pSL1180 to produce pSLC. The *Spe*I-*Sph*I and *Bal*I-*Eco*RV fragments from this plasmid pSLC, containing the 5'-end and the 3'-end of *borC* respectively, were then cloned stepwise into the *Spe*I-*Sph*I and *Eco*RV sites of pEFBA and in the correct orientations. In this way a 302bp *Sph*I-*Bal*I internal fragment of *borC* was replaced by the *aac(3)IV* gene. The resulting plasmid was then digested with *Spe*I and *Eco*RV and the resulting fragment was subcloned together with the *hyg* gene as described above, into pSL1180 leading to the final construct pSLCr1. This approach was used in order to avoid possible polar effects.

The vector pSLCr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borC::aac3(IV)*. Strain *S. parvulus* Tü4055/*borC::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was compared to a wild type control. In addition, *S. parvulus* Tü4055/*borC::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

To verify that no polar effects were introduced a full-length copy of *borC* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. Full-length *borC* was amplified by PCR using the primers B6T1 (5'-CGGATGCATCACCGGCACGG-3') (SEQ ID No.95) and B6T2 (5'-TGGGATCCGCGGGGCGGTAC-3') (SEQ ID No.96) using cosmid Bor32A2 as

template. The 943 bp product was purified and then digested with with *Nsi*I-*Bam*HI and subcloned, together with a *Bam*HI-*Spe*I fragment from pLHyg (carrying the *hyg* gene), into pIJ2925 previously digested with *Pst*I-*Xba*I. A *Bgl*II fragment (using this site from the vector) was then isolated and subcloned into pEM4, and in the correct orientation to locate *borC* under the control of the promoter ermE*. Plasmid pborCH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borC::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus* Tü4055/*borC::aac(3)IV*/pborCH was analysed as described in example 1 and shown to produce borrelidin at a titre similar to a WT control.

Example 15: Disruption of *borD* (*S. parvulus* Tü4055/*borD::aac3(IV)*)

In order to disrupt *borD*, a fragment of 2777 bp was amplified by PCR using the primers BBB (5'-AACTAGTGCGATCCCGGGGA-3') (SEQ ID No.97) and BBA (5'-CGTCGATATCCTCCAGGGGC-3') (SEQ ID No.98) and cosmid Bor32A2 as template. The PCR product was purified and then digested with *Spe*I-*Eco*RV and subcloned into pSL1180 to generate pSLD. This was then digested with *Nde*I-*Stu*I to delete an internal 679 bp region of *borD* which was replaced by a *Sma*I-*Nde*I fragment isolated from pEFBA containing the *aac(3)IV* gene. The resulting construct was digested with *Spe*I-*Eco*RV and the 4.3 kb fragment subcloned together with a *Spe*I-*Pst*I fragment from pLHyg containing the *hyg* gene, into pSL1180 digested with *Pst*I-*Eco*RV. This step leads to the final plasmid pSLDr1. This approach was used in order to avoid possible polar effects.

The vector pSLDr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borD::aac3(IV)*. Strain *S. parvulus* Tü4055/*borD::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was compared to a wild type control. In addition, *S. parvulus* Tü4055/*borD::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

To verify that no polar effects were introduced a full-length copy of *borD* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. Full-length *borD* was amplified by PCR using the primers BBT1 (5'-

TACTGCAGCACACCCGGTGC-3') (SEQ ID No.99) and BBT2 (5'-TGGGATCCGCTGTGTCATAT-3') (SEQ ID No.100) using cosmid Bor32A2 as template. The 816 bp PCR product was purified and then digested with with *Pst*I-*Bam*HI and subcloned together with a *Bam*HI-*Spe*I fragment containing the *hyg* gene from pLHyg, into pIJ2925 digested with *Pst*I-*Xba*I, to give pIJDH. The *Bgl*II fragment from pIJDH (using these sites from the vector) was then subcloned into pEM4 (predigested with *Bam*HI) and in the correct orientation to generate pborDH. Plasmid pborDH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borD::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus* Tü4055/*borD::aac(3)IV*/pborDH was analysed as described in example 1 and shown to produce borrelidin at a titre similar to a WT control.

Example 16: Disruption of *borE* (*S. parvulus* Tü4055/*borE::aac3(IV)*)

In order to disrupt *borE*, an internal 761 bp fragment of the gene was amplified by PCR using primers B25A (5'-TTCTGCAGCCGCGGCCTTCG-3') (SEQ ID No.81) and B25B (5'-AGAATTCGCCGGCGCCGCTG -3') (SEQ ID No.82) using cosBor32A2 as template. The product was purified, digested *Pst*I-*Eco*RI and cloned into pOJ260ermE* which had been digested similarly, to provide pOJEd1. This approach was used in order to avoid possible polar effects. The vector pOJEd1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5, and colonies were selected for apramycin resistance on R5 and then on MA agar. The disruption was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borE::aac3(IV)*. Strain *S. parvulus* Tü4055/ *borE::aac3(IV)* was grown, extracted and analysed as described in example 1. No borrelidin production was observed whereas a wild type control produced borrelidin as expected.

To verify that no polar effects were introduced a full-length copy of *borE* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. Full-length *borE* was amplified by PCR using the primers B7T1 (5'-GGCTGCAGACGCGGCTGAAG-3') (SEQ ID No.83) and B7T2 (5'-CCGGATCCCAGAGCCACGTC-3') (SEQ ID No.84) using cosBor32A2 as template. The 1216 bp product was purified, digested with *Pst*I-*Bam*HI and cloned into *Pst*I-*Xba*I digested pIJ2925 (Janssen & Bibb, 1993), along with a *Bam*HI-*Spe*I digested fragment from pLHyg containing the hygromycin resistance cassette, to generate pIJEH. A 2.8 kbp *Bam*HI fragment was excised from pIJEH and cloned into pEM4 (Quiros et al.,

1998), which had been digested similarly, to give pborEH (in which the *borE* gene was cloned in the correct orientation for gene expression). pborEH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borE::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus* Tü4055/*borE::aac(3)IV*/pborEH was analysed as described in example 1 and shown to produce borrelidin at a titre similar to a WT control; the control strain *S. parvulus* Tü4055/ *borE::aac(3)IV*/pEM4 did not produce borrelidin.

Chemical complementation of *S. parvulus* Tü4055/*borE::aac3(IV)* with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1, demonstrated that the strain thus grown was capable of borrelidin production at 122±23% of the WT parent control. Thus, *borE* is required for biosynthesis of *trans*-cyclopentane-1,2-dicarboxylic acid.

Example 17. Disruption of *borF* (*S. parvulus* Tü4055/*borF::aac3(IV)*)

In order to disrupt *borF*, a region containing *borF* was amplified by PCR using the primers BCB (5'-CACTAGTCCTCGCCGGGCAC-3') (SEQ ID No.101) and BCA (5'-GAGGATCCCGGTCAGCGGCA-3') (SEQ ID No.102) and cosmid Bor32A2 as template. The resulting 2132 bp product was purified and then digested with *Spe*I-*Bam*HI and subcloned into the same sites of pSL1180 leading to pSLF. The *aac(3)IV* gene from pEFBA was then subcloned as a *Sph*I fragment into the *Sph*I site of pSLF, which is located inside the *borF* coding region. Finally the *Bam*HI-*Spe*I fragment was subcloned into pLHyg digested with *Bam*HI-*Nhe*I to generate pLHFr1.

The vector pLHFr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borF::aac3(IV)*. Strain *S. parvulus* Tü4055/*borF::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was compared to a wild type control. In addition, *S. parvulus* Tü4055/*borF::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

To verify that no polar effects were introduced a full-length copy of *borF* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. Full-length *borF* was amplified by PCR using the primers BCT1 (5'-

GCCTGCAGCGACCTCGCCGG-3') (SEQ ID No.103) and BCT2 (5'-CGGGATCCCGTGGCGTGGTC-3') (SEQ ID No.104) using cosmid Bor32A2 as template. The 1048 bp PCR product was purified and then digested with *Pst*I-*Bam*HI and subcloned together with the *hyg* gene as described above, into pIJ2925. A *Bgl*II fragment was then isolated and subcloned into pEM4 to generate pborFH. This was used to complement strain SPMF. Plasmid pborFH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borF::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus* Tü4055/*borF::aac(3)IV*/pborFH was analysed as described in example 1 and shown to produce borrelidin at a titre similar to a WT control.

Example 18: Disruption of *borG* (*S. parvulus* Tü4055/*borG::aac3(IV)*)

In order to disrupt *borG*, an internal region of 885 bp was amplified by PCR using the primers B23A (5'-ATCTGCAGCGGCATCGGTGT-3') (SEQ ID No.105) and B23B (5'-AGAATTCTCCACTGCGGTCG-3') (SEQ ID No.106) and cosmid Bor32A2 as template. The resulting product was purified and the digested at the flanking sites *Pst*I-*Eco*RI and then subcloned into pOJ260P, downstream of the promoter *ermE**, to generate pOJGd1.

The vector pOJGd1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected on MA agar. The disruption was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borG::aac3(IV)*. Strain *S. parvulus* Tü4055/*borG::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was compared to a wild type control. In addition, *S. parvulus* Tü4055/*borG::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

Example 19: Disruption of *borH* (*S. parvulus* Tü4055/*borH::aac3(IV)*)

In order to disrupt borH, and internal region of 697 bp was amplified by PCR using the primers B9A (5'-ACCTGCAGGCCGGGCTCATC-3') (SEQ ID No.107) and B9B (5'-AGAATTCGGGCGAGCCGCCG-3') (SEQ ID No.108) and cosmid Bor32A2 as template. The resulting PCR product was purified and then digested with *Pst*I-*Eco*RI and then subcloned into pOJ260P, downstream of the promoter *ermE**, to generate pOJHd2.

The vector pOJHd2 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected on MA agar. The disruption was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borH::aac3(IV)*. Strain *S. parvulus* Tü4055/*borH::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was compared to a wild type control. In addition, *S. parvulus* Tü4055/*borH::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

Example 20: Disruption of *borI* (*S. parvulus* Tü4055/*borI::aac3(IV)*)

The gene *borI* and surrounding DNA was amplified from cosBor19B9 using the PCR primers BP4501 (5'-CGTATGCATGGCGCCATGGA-3') (SEQ ID No.85) and BP4502 (5'-AGCCAATTGGTGCACTCCAG-3') (SEQ ID No.86). The 2.32 kbp product was purified, digested with *Nsi*I-*Mfe*I and cloned into pSL1180 digested *Nsi*I-*Eco*RI, to give plasmid pSLI. The apramycin resistance cassette was excised from pEFBA as an *Eco*RI fragment and cloned into pSLI digested with *Eco*RI, to give the plasmid pSLIA. Finally, the hygromycin resistance cassette was excised *Spe*I-*Pst*I from pLHyg and cloned into pSLIA which had been digested with *Nsi*I-*Spe*I to give plasmid pSLIr1.

The replacement vector pSLIr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin (25 µg/ml) were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borI::aac3(IV)*.

S. parvulus Tü4055/*borI::aac3(IV)* was grown and analysed as described in example 1. No borrelidin production was observed whereas several new compounds were observed at significantly lower levels. One of the less polar compounds displayed a UV absorbance maximum of 240 nm, and LCMS analysis indicated an *m/z* ratio 11 mass units lower than that for borrelidin, which is consistent with the presence of a methyl- rather than a nitrile-group at C12.

To verify that no polar effects were introduced a full-length copy of *borI* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. A 2.1 kb *Nsi*I-*Avr*II fragment containing *borI* was recovered from pSLI and subcloned into the *Pst*I-*Xba*I sites of pEM4, together with the *Nhe*I-*Spe*I fragment from pLHyg containing the hyg gene. Both fragments were subcloned in the same orientation

generating pborIH. Plasmid pborIH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borI::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus* Tü4055/*borI::aac(3)IV*/pborIH was analysed as described in examples 1 & 4, and shown to produce borrelidin at a titre similar to a WT control.

<u>Example 21: Disruption of *borJ* (*S. parvulus* Tü4055/*borJ::aac3(IV)*)</u>

The gene *borJ* and surrounding DNA was amplified from cosBor19B9 using the PCR primers BNHT1 (5'-GTCATGCATCAGCGCACCCG-3') (SEQ ID No.87) and BNHT2 (5'-GTGCAATTGCCCTGGTAGTC-3') (SEQ ID No.88). The 2.75 kbp product was purified, digested with *Nsi*I-*Mfe*I and cloned into pSL1180 that had been digested with *Nsi*I-*Eco*RI, to give plasmid pSL. The hygromycin resistance cassette was excised from pLHyg as a *Pst*I-*Spe*I fragment and cloned into pSL digested with *Nsi*I-*Spe*I, to give pSLJH. Finally, the apramycin resistance cassette was excised from pEFBA with *Spe*I-*Bam*HI and cloned into pSLJH that had been pre-digested with *Avr*II-*Bgl*II in order to remove a 453 bp fragment from *borJ*, to give plasmid pSLJr1.

The replacement vector pSLJr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin (25 □g/ml) were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation. The new mutant was named *S. parvulus* Tü4055/*borJ::aac3(IV)*.

S. parvulus Tü4055/*borJ::aac3(IV)* was grown and analysed as described in example 1. No borrelidin production was observed whereas a new compound more polar than borrelidin was observed with a UV maximum at 262 nm. LCMS analysis indicated a parent compound of 508 amu, which is consistent with a carboxylic acid rather than a nitrile function at C12.

To verify that no polar effects were introduced a full-length copy of *borJ* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. A 2.4 kb *Nsi*I-*Sph*I fragment from pSLJ containing borJ was subcloned into the *Pst*I-*Xba*I sites of pEM4, together with the hyg gene as a *Sph*I-*Spe*I fragment from pLHyg; both fragments were subcloned in the same orientation as the transcription of the genes. The final construct was designed pborJH. Plasmid pborJH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borJ::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus*

Tü4055/*borJ::aac(3)IV*/pborJH was analysed as described in examples 1 & 4, and shown to produce borrelidin at a titre similar to a WT control.

Example 22: Disruption of *borK* (*S. parvulus* Tü4055/*borK::aac3(IV)*)

In order to disrupt *borK*, a fragment of 2680 bp was amplified by PCR using the primers B231 (5'-ATCAAGCTTCGTGTCCATGG-3') (SEQ ID No.109) and B232 (5'-GTCATGCATCAGGCGTTCGG-3') (SEQ ID No.110) and cosmid Bor19B9 as template. The resulting PCR product was purified and then digested with *Hind*III-*Nsi*I and subcloned into the same sites of pSL1180 to produce pSLK. After *Mlu*I digestion of pSLK and treatment with the Klenow fragment, the *aac(3)IV* gene from pEFBA was subcloned as a *Sma*I-*Eco*RV fragment leading to pSLKa. Finally a *Pst*I-*Spe*I fragment from pLHyg containing the *hyg* gene was subcloned into pSLKa digested *Nsi*I-*Xba*I to obtain pSLKr1.

The vector pSLKr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borK::aac3(IV)*. Strain *S. parvulus* Tü4055/*borK::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was compared to a wild type control. In addition, *S. parvulus* Tü4055/*borK::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

To verify that no polar effects were introduced a full-length copy of *borK* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. A 2.2 kb *Bgl*II (blunt-ended)-*Nsi*I fragment from pSLK was subcloned, together with a 1.6 kb *Pst*I-*Spe*I fragment from pLHyg containing the hyg gene, into pEM4 digested with *Pst*I (treated with the Klenow fragment) and then XbaI. The final vector was named pborKH. Plasmid pborKH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borK::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus* Tü4055/*borK::aac(3)IV*/pborKH was analysed as described in examples 1 & 4, and shown to produce borrelidin at a titre similar to a WT control.

Example 23: Disruption of *borL* (*S. parvulus* Tü4055/*borL::aac3(IV)*)

In order to disrupt *borL* a 3.95 kbp *Bgl*II fragment of cosBor19B9, which contained the full-length *borL*, was sub-cloned into pSL1180 digested similarly. The resulting clones were analysed by restriction digest and one that displayed the correct orientation was chosen to provide pSL395. Digestion of pSL395 with *Nhe*I and *Spe*I, and subsequent re-ligation to eliminate a fragment of *borM* that included a *Bgl*II site, gave pSLL. The apramycin resistance cassette was excised with *Kpn*I from pEFBA (Lozano *et al.*, 2000) and cloned into pSL that had been digested with *Kpn*I, to give pSLLA. pSLLA was digested with *Bgl*II and then subjected to Klenow treatment following the manufacturers instructions (Roche); an *Eco*RV fragment isolated from pLHyg containing the hygromycin resistance cassette was then cloned into this prepared vector to give pSLLr1.

The replacement vector pSLLr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation. Colonies resistant to apramycin were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation. The new mutant was named *S. parvulus* Tü4055/*borL::aac3(IV)*.

Strain *S. parvulus* Tü4055/*borL::aac3(IV)* was grown, extracted and analysed as described in example 1. No borrelidin production was observed whereas a wild type control produced borrelidin as expected. Chemical complementation of *S. parvulus* Tü4055/*borL::aac(IV)* using the natural starter acid as described in example 1 showed that the strain thus grown was capable of borrelidin production at 408±70 % of the WT parent control titre.

To verify that no polar effects were introduced a full-length copy of *borL* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. The vector containing full-length *borL* was generated as described in example 30. Plasmid pborLH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borL::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus* Tü4055/*borL::aac(3)IV*/pborLH was analysed as described in example 1.

Example 24: Disruption of *borM* (*S. parvulus* Tü4055/*borM::aac3(IV)*)

In order to disrupt *borM*, a 2870 bp fragment containing *borM* was amplified by PCR using the primers B251 (5'-CTTCTAGATGAACCCCTCCA-3') (SEQ ID No.111) and B252 (5'-GGGCAATTGCGCGGCAGCTT-3') (SEQ ID No.112) and cosmid

Bor19B9 as template. The resulting product was purified and then digested with *Xba*I-*Mfe*I and subcloned into the *Xba*I-*Eco*RI sites of pSL1180, leading to pSLM. An internal 780 bp *Sph*I-*Nhe*I fragment of *borM* was then replaced by the *aac(3)IV* gene which was subcloned from pEFBA as a *Spe*I-*Xba*I fragment, leading to pSLMA. pSLMA was digested with *Nsi*I-*Xba*I and the *hyg* gene subcloned as a *Spe*I fragment from pLHyg to generate pSLMr1.

The vector pSLMr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borM::aac3(IV)*. Strain *S. parvulus* Tü4055/*borM::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was compared to a wild type control. In addition, *S. parvulus* Tü4055/*borM::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

To verify that no polar effects were introduced a full-length copy of *borM* under the control of the ermE* promoter was introduced *in trans* to the disrupted mutant. Full-length *borM* was cloned as a *Xba*I-*Age*I fragment of 2.0 kb from pSLM and subcloned into the *Eco*RI (end-filled with Klenow)-*Xba*I sites of pEM4 together with the *hyg* gene as a *Xma*I-*Eco*RV fragment from pLHyg, to give pborMH. Plasmid pborMH and the control plasmid pEM4 were introduced into *S. parvulus* Tü4055/*borM::aac(3)IV* by protoplast transformation as described in example 5. The resulting strain *S. parvulus* Tü4055/*borM::aac(3)IV*/pborMH was analysed as described in example 1 and shown to produce borrelidin at a titre similar to a WT control.

Example 25: Disruption of *borN* (*S. parvulus* Tü4055/*borN::aac3(IV)*)

In order to disrupt borN, a 1201 bp *Bam*HI fragment from pSLM (containing the 3'-end of *borM* and the first 161 codons of *borN*) was subcloned into the *Bgl*II-*Bam*HI sites of pSL1180 and in the correct orientation, to generate pSLMN. A *Bam*HI-*Eco*RI fragment (using these sites from the polylinker) containing *borO* from pborOR (see below) was subcloned into the *Bam*HI-*Eco*RI sites of pSLMN, generating pSLNO. After *Eco*RI digestion of pSLNO and end-filling with Klenow fragment, the *hyg* gene was subcloned from pLHyg as a *Eco*RV fragment, leading to pSLNOH. Finally the *aac3(IV)*

gene was subcloned as a *Nco*I-*Bam*HI fragment from pEFBA into pSLNOH digested with the same restriction enzymes, generating pSLNr1.

The vector pSLNr1 was introduced into *S. parvulus* Tü4055 by protoplast transformation as described in example 5. Colonies resistant to apramycin were selected, and then passaged several times through MA media without selection. The replacement was verified by Southern hybridisation and the new mutant was named *S. parvulus* Tü4055/*borN::aac3(IV)*. Strain *S. parvulus* Tü4055/*borN::aac3(IV)* was grown, extracted and analysed as described in example 1. Borrelidin production was compared to a wild type control. In addition, *S. parvulus* Tü4055/*borN::aac3(IV)* was chemically complemented with *trans*-1,2-dicyclopentane dicarboxylic acid, following the protocol described in example 1.

<u>Example 26: Heterologous expression of *borO* in *Streptomyces albus* J1074</u>

In order to examine whether the putative resistance protein BorO confers resistance to a borrelidin-sensitive organism, *borO* was expressed in *Streptomyces albus* J1074. The gene *borO* was amplified by PCR using the primers BTRNAS1 (5'-TGTCTAGACTCGCGCGAACA-3') (SEQ ID No.89) and BTRNAS2 (5'-TGAATTCCGAAGGGGGTGGT-3') (SEQ ID No.90) with cosBor19B9 as template. The product was purified, digested *Xba*I-*Eco*RI and cloned into pEM4A that had been similarly digested to give plasmid pborOR which puts *borO* under the control of the promoter ermE*. The vector pborOR was introduced into *S. albus* J1074 by protoplast transformation (Chater & Wilde, 1980) and selected for apramycin resistance. The new strain was named *S. albus* J1074/pborOR.

Resistance to borrelidin was assayed on Bennett's agar containing apramycin at 25 µg/ml. Spores of *S. albus* J1074/pborOR and the control *S. albus* J1074/pEM4A were spread onto plates and then disks containing borrelidin at 100 & 200 µg/ml were laid upon the lawn of spores and incubated overnight at 30°C. Haloes indicating inhibition of growth were observed for the control strain harbouring pEM4A but not for *S. albus* J1074/ pborOR.

<u>Example 27: Disruption of *borG* and *borI* (*S. parvulus* Tü4055/*borG::aac3(IV)*/*borI::hyg*)</u>

The hyg gene is isolated from pLHyg as an *Eco*RV fragment and cloned into pSLI (example 20) digested with *Eco*RI and treated with Klenow fragment to give pSLIH; the hyg gene is cloned in the same orientation as *borI*. pSLIH is introduced into

S. parvulus Tü4055/*borG::aac3(IV)* by protoplast transformation, as described in example 5, and selected for both apramycin and hygromycin resistance, and is then passaged several times through MA media without selection in order to promote double recombination. Apramycin and hygromycin resistant colonies are analysed by Southern hybridisation and PCR to verify the replacement.

Example 28: Disruption of *borG* and *borJ* (*S. parvulus* Tü4055/*borG::aac3(IV)*/*borJ::hyg*)

The hyg gene is isolated from pLHyg as an *Eco*RV fragment and cloned into pSLJ (example 21) digested with *Avr*II-*Bgl*III and treated with Klenow, to give pSLJH; the hyg gene is cloned in the same orientation as *borI*. pSLJH is introduced into *S. parvulus* Tü4055/*borG::aac3(IV)* by protoplast transformation, as described in example 5, and selected for both apramycin and hygromycin resistance, and is then passaged several times through MA media without selection in order to promote double recombination. Apramycin and hygromycin resistant colonies are analysed by Southern hybridisation and PCR to verify the replacement.

Example 29: Effects of *borE* up-regulation in *S. parvulus* Tü4055

To examine the possibility that biosynthesis of the *trans*-1,2-cyclopentane dicarboxylic acid starter unit may have a limiting effect upon borrelidin production, *borE* was up-regulated in the parental strain and the effect upon borrelidin titre was analysed. The vector used, pborEH was described in example 16.

The vectors pborEH and pEM4 (control) were used to transform protoplasts of *S. parvulus* Tü4055 to give strains *S. parvulus* Tü4055/pborEH and *S. parvulus* Tü4055/pEM4 respectively. Several colonies from each transformation were picked, grown in triplicate and then analysed as described in example 1. Compared to the control strain, up-regulation of *borE* brought about a 4.2±0.3-fold increase in the titre of borrelidin.

Example 30: Effects of *borL* up-regulation in *S. parvulus* Tü4055

To examine the possibility that *borL* may have a regulatory, or some other related function involved in borrelidin production, the gene was up-regulated in the parental strain and the effect upon borrelidin titre was analysed.

The expression vector pborLH was generated as follows: pSLL was digested with *Not*I, treated with Klenow fragment and then digested with *Bam*HI to obtain a fragment of 2190 bp containing *borL*. This fragment was sub-cloned together with the *Bam*HI-*Spe*I *hyg* gene from pLHyg, into pEM4 digested with *Pst*I(treated with Klenow)-*Xba*I, to obtain pborLH.

The vectors pborLH and pEM4 (control) were used to transform protoplasts of *S. parvulus* Tü4055 to give strains *S. parvulus* Tü4055/pborLH and *S. parvulus* Tü4055/pEM4 respectively. Several colonies from each transformation were picked, grown in triplicate and then analysed as described in example 1. Compared to the control strain, up-regulation of *borL* brought about a 4.3±0.7-fold increase in the titre of borrelidin.

Example 31: Production of 12-desnitrile-12-methyl borrelidin **14** (pre-borrelidin)

Working stocks of *S. parvulus* Tü4055/*borI::aac3(IV)* (0.5 ml) were inoculated into primary vegetative pre-cultures of NYG as described in example 1. Secondary pre-cultures were prepared (as example 1 but with 250 ml NYG in 2 l Erlenmeyer flasks). PYDG production medium (4 l), prepared as in example 1 and with 0.01 % Plutronic L0101 added to control foaming, was inoculated with secondary pre-culture (12.5 % inoculum). A second fermenter containing centre-point medium (4 l) and 0.01 % Plutronic L0101 to control foaming, was set up in parallel and was also inoculated with secondary pre-culture (12.5 % inoculum). Centre-point production medium contains per litre of deionised water: Tesco's skimmed milk powder (1.5 %), Avidex W-80 (4.5 %), glucose (0.5 %) and yeast autolysate (0.15 %) adjusted to pH 7.0 with 5 M NaOH.

These batches were each allowed to ferment in a 7 l Applikon fermenter for 6.5 days at 30 °C. Airflow was set at 0.75 vvm (volume per volume per minute), with tilted baffles and the impeller speed controlled between 400 and 800 rpm to maintain dissolved oxygen tension at or above 30 % of air saturation. No further antifoam was added. At 22 hours into the fermentation the starter acid, *trans*-cyclopentane-1,2-dicarboxylic acid, was added as a neutralised solution of 1:1 MeOH / 5 M NaOH, through an in-line filter (0.22 μm). The final concentration in the fermenter vessel of exogenous starter acid was 0.5 mM.

After 6.5 days of fermentation the broths were combined and acidified to pH 3.5 with concentrated HCl (~ 6 ml), then clarified by centrifugation at 3,500 rpm for 10 minutes. The supernatant was extracted into ethyl acetate (3 x 1 volume equivalent for

4 hours each) and the cell pellet left to steep in methanol (2 x 1.5 litres for 4 hours each). The organics were combined and removed under reduced pressure to yield a tarry gum. The gum was re-suspended in 0.1 M Borax buffer (500 ml at pH 9.4) and washed with hexanes (500 ml) and ethyl acetate (500 ml). The aqueous layer was then acidified with concentrated HCl to pH 3.5 and extracted with ethyl acetate (3 x 500 ml), which were combined and taken to dryness. The resultant gum was dissolved in methanol (15 ml), diluted with water (285 ml) and loaded under gravity onto a C_{18}-reversed-phase cartridge (50 g, prepared in 5 % aqueous methanol). The cartridge was washed with 20 % and 50 % aqueous methanol (300 ml each) and eluted with 100 % methanol (500 ml). This last fraction was taken to dryness under reduced pressure to yield a black gummy-oil (600 mg) that was taken up in methanol. This residue was finally purified by sequential preparative reversed-phase HPLC (eluted with the mobile phases used in example 4, without added TFA, running isocratically at 40 %B). Active fractions were combined and desalted on a C_{18}-cartridge (1 g), to yield 28 mg of a dark oil (3.5 mg/l isolated yield). Table 12 summarises the ^{1}H and ^{13}C NMR chemical shift data for 12-desnitrile-12-methyl borrelidin **14** in $CDCl_3$.

Table 12

Position	δ_H (ppm)	Multiplicity	Coupling (Hz)	δ_C (ppm)
1	-	-	-	174.5
2a	2.29	m	-	37.8
2b	2.26	m	-	-
3	3.85	dt	9.0, 3.0	71.9
4	1.83	m	-	35.1
5a	1.19	bt	13.5	43.6
5b	0.91	m	-	-
6	1.75	m	-	27.0
7a	1.08	m	-	49.2
7b	0.88	m	-	-
8	1.69	m	-	26.5
9a	0.97	m	-	38.3
9b	0.45	t	12.5	-
10	1.62	m	-	34.1
11	3.53	d	9.0	85.7

Position	δ_H (ppm)	Multiplicity	Coupling (Hz)	δ_C (ppm)
12	-	-	-	138.4
13	5.84	d	11.0	127.7
14	6.28	ddd	14.5, 11.0, 1.0	129.6
15	5.48	ddd	14.5, 10.5, 3.5	129.9
16a	2.53	m	-	39.1
16b	2.22	m	-	-
17	5.07	ddd	11.0, 8.0, 3.0	76.5
18	2.52	m	-	48.0
19a	1.92	m	-	30.4
19b	1.32	m	-	-
20a	1.74	m	-	26.2
20b	1.71	m	-	-
21a	1.96	m	-	32.0
21b	1.84	m	-	-
22	2.45	m	8.0	49.3
23	-	-	-	182.3
4-CH_3	0.78	d	6.5	18.5
6-CH_3	0.77	d	6.5	18.8
8-CH_3	0.75	d	6.5	20.6
10-CH_3	0.94	d	6.5	16.3
12-CH_3	1.64	s	-	11.4

Chemical shifts are referenced to $CDCl_3$ (for ^1H at 7.26 ppm and for ^{13}C at 77.0 ppm)

Example 32: Production of 12-desnitrile-12-carboxy borrelidin **2**

Working stocks of *S. parvulus* Tü4055/*borJ::aac3(IV)* (0.5 ml) were inoculated into primary vegetative pre-cultures of NYG as described in example 1. Secondary pre-cultures were prepared (as example 1 but with 250 ml NYG in 2 l Erlenmeyer flasks). PYDG production media (4 L), prepared as in example 1 and with 0.01 % Plutronic L0101 added to control foaming, was inoculated with the entire secondary pre-culture (10 % inoculum). This was allowed to ferment in a 7 L Applikon fermenter for 6 days at 30 °C. Airflow was set at 0.75 vvm, with tilted baffles and the impeller speed controlled between 250 and 600 rpm to maintain dissolved oxygen tension at or above 30 % of air

saturation. No further antifoam was added. A second fermentation was performed exactly as above, but which was batch fed with 0.2 mol of glucose as an aqueous solution every 12 hours from 60 hours post-inoculation.

After 6 days the fermentations were harvested and combined. The broth was clarified by centrifugation (3,500 rpm, 10 minutes) and the resultant supernatant acidified with 10 M HCl (aq) to pH ~ 3.5. This solution was then extracted into ethyl acetate by stirring (3 x 1 volume equivalent for 4 hours each). The cell pellet was extracted twice by steeping the cells in 1:1 methanol / ethyl acetate (500 ml). All the organics were combined and removed under reduced pressure to yield an aqueous slurry. The slurry was diluted to 500 ml with water, acidified to pH ~ 3.5 with 10 M HCl and extracted into ethyl acetate (3 x 300 ml). The organics were concentrated under reduced pressure to ~ 300 ml and extracted with 0.1 M borax (3 x 150 ml, pH = 9.4). The combined borax solutions were acidified with 10 M HCl to pH ~ 3.5 and extracted with 6 x 300 ml of ethyl acetate. Analytical HPLC demonstrated that some of the accumulant still resided in the borax solution and so this was loaded, under gravity, onto a C_{18}-reverse-phase cartridge (50 g). The cartridge was washed with water and the accumulant eluted in 100 % methanol. The organics containing the accumulant were combined and reduced to a 40 ml methanolic solution. This was loaded onto a Sephadex LH-20 column (70 g, swelled overnight in methanol, column 60 cm x 2.5 cm), which was developed with 100 % methanol; the active fractions were combined and taken to dryness. The material was then further processed by preparative reversed-phase HPLC (eluted with the mobile phases used in example 4, without added TFA, running isocratically at 40 %B). The combined active fractions were taken to dryness, dissolved in methanol (4 ml) and diluted with water (200 ml). This mixture was split into 2 equal fractions and each loaded, under gravity, onto a C_{18}-reverse-phase cartridge (20 g). The columns were then eluted with 3 column volumes of 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 75 %, 90 % and 100 % aqueous methanol. The accumulant eluted in all fractions from 60 % to 100 % methanol, which were combined and taken to dryness. The accumulant (dissolved in DMSO) was then finally purified by sequential preparative reversed-phase HPLC (eluted with the mobile phases used in example 4, without added TFA, running isocratically at 40 %B). Active fractions were combined and desalted on a C_{18}-cartridge (1 g), to yield 17 mg of a brown oil (2.1 mg/l isolated yield). Table 13 summarises the ^{1}H and ^{13}C NMR chemical shift data for 12-desnitrile-12-carboxy borrelidin **2** in d_4-methanol.

Table 13

Position	δ_H (ppm)	Multiplicity	Coupling (Hz)	δ_C (ppm)
1	-	-	-	173.27
2a	2.40	dd	15.8, 4.1	39.31
2b	2.29	dd	15.8, 8.2	
3	3.87	m		71.64
4	1.80	m		36.51
5a	1.29	m		44.24
5b	0.90	m		
6	1.59	m		27.48
7a	1.09	m		~49.0*
7b	1.03	m		
8	1.72	m		28.17
9a	1.12	m		38.42
9b	0.79	m		
10	2.03	m		36.43
11	3.90	m		81.95
12	-	-	-	132.35
13	6.43	d	11.0	140.83
14	6.96	dd	14.5, 11.5	130.91
15	5.91	ddd	15.0, 9.5, 5.0	138.93
16a	2.61	m	15.0	38.57
16b	2.36	m		
17	5.04	m		77.40
18	2.50	m		49.80
19a	1.90	m		30.59
19b	1.32	m		
20a	1.85	m		26.34
20b	1.41	m		
21a	1.97	m		32.40
21b	1.75	m		
22	2.52	m		~48.0*

Position	δ_H (ppm)	Multiplicity	Coupling (Hz)	δ_C (ppm)
23	-	-	-	180.27
4-CH_3	0.83	d	7.0	18.76
6-CH_3	0.80	d	6.0	17.06
8-CH_3	0.81	d	6.5	20.60
10-CH_3	0.93	d	6.5	16.61
12-CO_2H	-	-	-	170.49

Chemical shifts are referenced to methanol (for ^{1}H at 3.35 ppm (quintet) and for ^{13}C at 49.0 ppm (septet)); *Obscured by solvent signal, d_4-methanol.

Example 33: Production by mutasynthesis of 17-des-(cyclopentane-2'-carboxylic acid)-17-(cyclobutane-2'-carboxylic acid)borrelidin **18**

Working stocks of *S. parvulus* Tü4055/*borE::aac3(IV)* (0.5 ml) were inoculated into primary vegetative pre-cultures of NYG as described in example 1. Secondary pre-cultures were prepared (as example 1 but with 250 ml NYG in 2 l Erlenmeyer flasks). PYDG production medium (4 l), prepared as in example 1 and with 0.01 % Plutronic L0101 added to control foaming, was inoculated with secondary pre-culture (12.5 % inoculum). Two further bioreactors were set up in the same manner. These batches were each allowed to ferment in a 7 l Applikon fermenter for 5 days at 30 °C. Airflow was set at 0.75 vvm (volume per volume per minute), with tilted baffles and the impeller speed controlled between 400 and 700 rpm to maintain dissolved oxygen tension at or above 30 % of air saturation. No further antifoam was added. At 22 hours into the fermentation the starter acid, *trans*-cyclobutane-1,2-dicarboxylic acid, was added as a neutralised solution of 1:1 MeOH / 5 M NaOH. The final concentration in the fermenter vessel of exogenous starter acid was 0.5 mM.

After 5 days of fermentation the broths were combined and acidified to pH 4.0 with concentrated HCl, then clarified by centrifugation at 3,500 rpm for 10 minutes. The supernatant was absorbed onto diaion HP-20SS resin (1 l), which had been pretreated with methanol (2 l) and then 5 % aqueous methanol (2 l), by filtration at a rate of approximately 100 ml/min. The resin was then eluted with 20 % aqueous methanol (2.5 l) and then 80 % aqueous acetone (4.5 l). The organic solvent was removed from the aqueous acetone and the resultant aqueous slurry (1 litre) extracted into ethyl acetate (3 x 1 l). The organics were combined and reduced *in vacuo* to yield a yellow/brown oil (1.7 g). Meanwhile, the cell pellet left to steep in methanol-ethyl acetate, 1:1 (3 x 1 l for

4 hours each), and the resultant organic supernatants reduced in vacuo to yield an aqueous slurry (400 ml). The particulate matter was dissolved in methanol (50 ml), and added back to the aqueous slurry, which was made up to 500 ml with water. This slurry was absorbed onto diaion HP-20SS resin (300 ml), that had been pretreated with methanol (500 ml) and then 5 % aqueous methanol (500 ml). The resin was then eluted with 20 % aqueous methanol (1 l) and then 80 % aqueous acetone (1.5 l). The organic solvent was removed from the aqueous acetone and the resultant aqueous slurry (made up to 750 ml) extracted into ethyl acetate (3 x 750 ml). The organics were combined and reduced *in vacuo* to yield a yellow/brown oil (1.7 g). The crude extracts were combined (3.4 g), dissolved in ethyl acetate (10 ml), then adsorbed onto a silica column (5 cm ID x 10 cm, treated with EtOAc), and eluted with EtOAc. The active fractions were combined and the solvent removed *in vacuo* to yield a brown gum (1.08 g). This residue was finally purified by sequential preparative reversed-phase HPLC (eluted with the mobile phases used in example 4, without added TFA, running from 25 % B to 75 % B over 25 minutes with a linear gradient). Active fractions were combined and desalted on a C_{18}-cartridge (5 g), to yield 83.9 mg (or 7.0 mg / l isolated yield). The ^{13}C-NMR spectrum of **18** is shown in table 14

Table14:

δ_c (ppm)	Position	δ_c (ppm)	Position
177.1	COOH (C22)	37.3	9
172.2	1	35.7	4
144.0	13	35.1	10
138.7	15	34.4	16
126.9	14	30.9	18
118.3	12	27.3	6
115.8	CN	26.2	8
75.5	17	21.7	20
73.1	11	21.0	21
69.7	3	20.1	8-Me
47.6	5	18.1	6-Me
43.1	7	16.9	4-Me
40.1	19	14.9	10-Me
40.0	2		

^{13}C-NMR assignment for **18**, in $CDCl_3$, using that carbon signal as reference at δ_C = 77.7 ppm

REFERENCES

Anderson, B.F., Herlt, A.J., Rickards, R.W., and Robertson, G.B. (1989) Crystal and molecular structures of two isomorphous solvates of the macrolide antibiotic borrelidin: absolute configuration determination by incorporation of a chiral solvent in the in the crystal lattice. *Aust. J. Chem.* **42**:717-730.

Anderton, K., and Rickards, R.W. (1965) Some structural features of borrelidin, an anti-viral antibiotic. *Nature* **206**:269.

Aparicio, J.F., Molnár, I., König, A., Haydock, S.H., Khaw., L.E., Staunton, J., and Leadlay, P.F. (1996) Organisation of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169**:9-16.

August, P.R., Tang, L., Yoon, Y.J., Ning, S., Müller, R., Yu, T.-W., Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X.H., Hutchinson, C.R., and Floss, H.G. (1998) Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem. Biol.* **5**:69-79.

Beck, J.B., Yoon, Y.J., Reynolds, K.A., and Sherman, D.H. (2002) The hidden steps of domain skipping: ring size determination in the pikromycin modular polyketide synthase. *Chem. Biol.* **9**:575-583.

Berger, J., Jampolsky, L.M., and Goldberg, M.W. (1949) Borrelidin, a new antibiotic with anti-Borrelia activity and penicillin enhancement properties. *Arc. Biochem.* **22**:476-478.

Bierman, M., Logan, R., O'Brian, K., Seno, E.T., Rao, N., and Schoner, B.E. (1992) Plasmid vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**:43-49.

Bonfield, J.K., Smith, K.F., and Staden, R. (1995) A new DNA sequence assembly program. *Nucleic Acids Research* **23**:4992-4999.

Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strøm, A.R., Valla, S., and Zotchev, S.B. (2000) Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. *Chem. Biol.* **7**:395-403.

Brenner, S. (1998) The molecular evolution of genes and proteins: a tale of two serines. *Nature* **334**:528-530.

Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F., and Weissman, K.J. (2003) The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**:723-731.

Brosius, J. (1989) Super-polylinkers in cloning and expression vectors. *DNA* **8**:759-777.

Butler, A.R., Bate, N., and Cundliffe, E. (1999) Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*. *Chem. Biol.* **6**:287-292.

Caffrey, P., Lynch, S., Flood, E., Finnan, S., and Oliynyk, M. (2001) Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. *Chem. Biol.* **8**:713-723.

Celenza, J.L. (2001) Metabolism of tyrosine and tryptophan – new genes for old pathways. *Curr. Opin. Plant Biol.* **4**:234-240

Chater, K.F. and Wilde, L.C. (1980) *Streptomyces albus* G mutants defective in the SalG1 restriction modification system. *J. Gen. Microbiol.* **116**:323-334.

Cheng, Y.Q., Tang, G.L., and Shen B. (2003) Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. USA.* **100**:3149-3154.

Cortés J., Haydock, S.F., Roberts, G.A., Bevitt, D.J., and Leadlay, P.F. (1990) An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**:176-178.

Cortés, J., Weissman, K.E.H., Roberts, G.A., Brown, M.J.B., Staunton, J., and Leadlay, P.F. (1995) Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268**:1487-1489.

Cortés, J., Velasco, J., Foster, G., Blackaby, A.P., Rudd, B.A.M., and Wilkinson, B. (2002) Identification and cloning of a type III polyketide synthase required for diffusible pigment biosynthesis in *Saccharopolyspora erythraea*. *Mol. Micro.* **44**:1213-1224.

Devereux, J., Heaberli, P., and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**:387-395.

Dickinson, L., Griffiths, A.J., Mason, C.G., and Mills, R.F. (1965) Anti-viral activity of two antibiotics isolated from a species of *Streptomyces*. *Nature* **206**:265-268.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**:675-679.

Donadio, S., McAlpine, J.B., Sheldon, P.J., Jackson, M., and Katz, L. (1993) An erythromycin analog produced by reprogramming of polyketide synthesis *Proc. Nat. Acad. Sci. USA* **90**:7119-7123.

Duffey, M.O., LeTiran, A., and Morken, J.P. (2003) Enantioselective total synthesis of borrelidin. *J. Am. Chem. Soc.* **125**:1458-1459.

Eastwood, E.L., and Schaus, S.E. (2003) Borrelidin induces the transcription of amino acid biosynthetic enzymes via a *GCN4*-dependent pathway. *Bioorg. Med. Chem. Lett.* **13**:2235-2237.

Fernandez, E., Weissbach, U., Sanchez-Reillo, C., Brana, A.F., Mendez, C., Rohr, J., and Salas, J.A. (1998) Identification of two genes from *Streptomyces argillaceus* encoding glycosyltransferases involved in transfer of a disaccharide during the biosynthesis of the antitumor drug mithramycin. *J. Bacteriol.* **180**:4929-4937.

Floss, H.G. (2001) Antibiotic biosynthesis: from natural to unnatural compounds. *J. Ind. Micro. Biotech.* **27**:183-194.

Fouces, R., Mellado, E., Diez, B., and Barredo, J.L. (1999) The tylosin biosynthetic cluster from *Streptomyces fradiae*: genetic organisation of the left region. *Microbiology* **145**:855-868.

Folkman, J. (1986) How is blood vessel growth regulated in normal and neoplastic tissue? G.H.A. Cloves Memorial Lecture. *Cancer Res.* **51**:467-473.

Funahashi, Y., Wakabayashi, T., Semba, T., Sonoda, J., Kitoh, K., and Yoshimatsu, K. (1999) Establishment of a quantitative mouse dorsal air sac model and its application to evaluate a new angiogenesis inhibitor. *Oncol. Res.* **11**:319-329.

Gaisser, S., Reather, J., Wirtz, G., Kellenberger, L., Staunton, J., and Leadlay, P.F. (2000) A defined system for hybrid macrolide biosynthesis in *Saccharopolyspora erythraea*. *Mol. Microbiol.* **36**:391-401.

Gaisser, S., Martin, C.J., Wilkinson, B., Sheridan, R.M., Lill, R.E., Weston, A.J., Ready, S.J., Waldron, C., Crouse, G.C., Leadlay, P.F., and Staunton, J. (2002) Engineered biosynthesis of novel spinosyns bearing altered deoxyhexose substituents. *Chem. Commun.* 618-619.

Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blöcker, H., Höfle, G., and Müller, R. (2002) The biosynthesis of the aromatic myxobacterial electron transport inhibitor

stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem.* **277**:13082-13090.

Hanessian, S., Yang, Y., Giroux, S., Mascitti, V., Ma, J., and Raeppel, F. (2003) Application of conformation design in acyclic stereoselection: total synthesis of borrelidin as the crystalline benzene solvate. *J. Am. Chem. Soc.* **125**:13784-13792.

Hardt, I.H., Steinmetz, H., Gerth, K., Sassa, F., Reichenbach, H., and Höfle, G. (2001) New natural epothilones from *Sorangium cellulosum*, strains So ce90/B2 and So ce90/D13: isolation, structure elucidation, and SAR studies. *J. Nat. Prod.* **64**:847-856.

Heathcote, M.L., Staunton, J., and Leadlay, P.F. (2001) Role of type II thioesterases: evidence for the removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**:207-220.

Hopwood, D. (1997) Genetic contributions to understanding polyketide biosynthesis. *Chem. Rev.* **97**:2465-2497.

Hunziker, D., Yu, T.-W., Hutchinson, C.R., Floss, H.G., and Khosla, C. (1998) Primer unit specificity in biosynthesis principally resides in the later stages of the biosynthetic pathways. *J. Am. Chem. Soc.* **120**:1092-1093.

Janssen, G.R., Bibb, M.J., (1993) Derivatives of pUC18 that have *Bgl*II sites flanking a modified cloning site and that retain the ability to identify recombinant clones by visual screening of *E. coli* colonies. *Gene* **124**:133-134.

Kahn, R.A., Fahrendorf, T., Halkier, B.A., and Møller, B.L. (1999) Substrate specificity of the cytochrome P450 enzymes CYP79A1 and CYP71E1 involved in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolour* (L.) Moench. *Arch. Biochem. Biophys.* **363**:9-18.

Kawamura, T., Liu, D., Towle, M.J., Kageyama, R., Tsukahara, N., Wakabayashi, T., and Littlefield, B.A. (2003) Anti-angiogenesis effects of borrelidin are mediated through

distinct pathways: Threonyl-tRNA synthetase and caspases are independently involved in suppression of proliferation and induction of apoptosis in endothelial cells. *J. Antibiot.* **56**:709-715.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000) Practical *Streptomyces* Genetics. The John Innes Foundation. Norwich.

Keller-Scheirlein, W. (1967) Composition of the antibiotic borrelidin. *Helv. Chim. Acta.* **50**:731-753.

Kuo, M.S., Yurek, D.A., and Kloosterman, D.A. (1989) Assignment of ^{1}H and ^{13}C NMR signals and the alkene geometry at C-7 in borrelidin. *J. Antibiot.* **42**:1006-1007.

Kuhstoss, S., Huber, M., Turner, J. R., Paschal, J. W., and Rao, R. N. (1996) Production of a novel polyketide through the construction of a hybrid polyketide synthase. *Gene* **183**:231-236.

Lozano, M.J., Remsing, L.L., Quiros, L.M., Brana, A.F., Fernandez, E., Sanchez, C., Mendez, C., Rohr, J., and Salas, J.A. (2000) Characterization of two polyketide methyltransferases involved in the biosynthesis of the antitumor drug mithramycin by *Streptomyces argillaceus*. *J. Biol. Chem.* **275**:3065-3074.

Maehr, H., and Evans, R.H. (1987) Identity of borrelidin with treponemycin. *J. Antibiot.* **40**:1455-1456.

Marsden, A.F., Wilkinson, B., Cortés, J., Dunster, N.J., Staunton, J., and Leadlay, P.F. (1998) Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science* **279**:199-202.

Matter, A,. (2001) Tumor angiogenesis as a therapeutic target. *Drug Dis. Today* **6**:1005-1024.

Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K., and Kinashi, H. (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol Microbiol.* **48**:1501-1510.

Moore, B.S., and Hopke, J.N. (2000) Discovery of a new bacterial polyketide biosynthetic pathway. *Chembiochem* **2**:35-38.

Nielsen, J.S., and Møller, B.L. (1999) Biosynthesis of cyanogenic glucosides in *Triglochin maritime* and the involvement of cytochrome P450 enzymes. *Arch. Biochem. Biophys.* **368**:121-130.

Olano, C., Wilkinson, B., Moss, S.J., Brana, A.F., Mendez, C., Leadlay, P.F., and Sala, J.A. (2003) Evidence from engineered gene fusions for the repeated use of a module in a modular polyketide synthase. *Chem. Commun.* 2780-2782.

Oliynyk, M., Brown, M.J.B., Cortés, J., Staunton., J., and Leadlay, P.F. (1996) *Chem. Biol.* **3**:833-839.

Otani, A., Slike, B.M., Dorrell, H.I., Hood, J., Kinder, K., Cheresh, D.A., Schimmel, P., and Friedlander, M. (2002) A fragment of human TrpRS as a potent antagonist of ocular angiogenesis. *Proc. Nat. Acad. Sci. USA* **99**:178-183.

Otoguru, K., Ui, H., Ishiyama, A., Kobayashi, M., Togashi, H., Takahashi, Y., Masuma, R., Tanaka, H., Tomado, H., Yamada, H., and Ōmura, S. (2003) *In vitro* and *in vivo* antimalarial activities of a non-glycosidic 18-membered macrolide antibiotic, borrelidin, against drug-resistant strains of *Plasmodia*. *J. Antibiot.* **56**:727-729.

Pacey, M.S., Dirlam, J.P., Geldart, L.W., Leadlay, P.F., McArthur, H.A.I., McCormick, E.L., Monday, R.A., O'Connell, T.N., Staunton, J., and Winchester, T.J. (1998) Novel erythromycins from a recombinant *Saccharopolyspora erythraea* strain NRRL 2338 pIG1 I. Fermentation, isolation and biological activity. *J. Antibiot.* **51**:1029-1034.

Paetz, W., and Nass, G. (1973) Biochemical and immunological characterization of threonyl-tRNA synthetase of two borrelidin-resistant mutants of *Escherichia coli* K12. *Eur. J. Biochem.* **35**:331-337.

Prieto, M.A., Díaz, E., and García, J.L. (1996) Molecular characterization of the 4-hydroxyphenylacetate catabolic pathway of *Escherichia coli* W: engineering a mobile aromatic degradative cluster. *J. Bacteriol.* **178**:111-120.

Quiros, L.M., Aguirrezabalaga, I., Olano, C., Mandez, C., and Salas, J.A. (1998) Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by *Streptomyces antibioticus. Mol. Microbiol.* **28**:1177-1185.

Raibaud, A., Zalacain, M., Holt, T.G., Tizard, R., and Thompson, C.J. (1991) Nucleotide sequence analysis reveals linked N-acetyl hydrolase, thioesterase, transport, and regulatory genes encoded by the bialophos biosynthetic gene cluster of *Streptomyces hygroscopicus. J. Bacteriol.* **173**:4454-4463.

Ranganathan, A., Timoney, M., Bycroft, M., Cortés, J., Thomas, I.P., Wilkinson, B., Kellenberger, L., Hanefeld, U., Galloway, I.S., Staunton, J., and Leadlay, P.F. (1999) Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: a route to simple statin analogues. *Chem. Biol.* **6**:731-741.

Reeves, C.D., Murli, S., Ashley, G.W., Piagentini, M., Hutchinson, C.R., and McDaniel, R. (2001) Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* **40**:15464-15470.

Rowe, C.J., Böhm, I.U., Thomas, I.P., Wilkinson, B., Rudd, B.A.M., Foster, G., Blackaby, A.P., Sidebottom, P.J., Roddis, Y., Buss, A.D., (2001) *Chem. Biol.* **8**:475-485.

Rowe, C.J., Cortés, J., Gaisser, S., Staunton, J., and Leadlay, P.F. (1998) Construction of new vectors for high-level expression in actinomycetes. *Gene* **216**:215-223.

Rudd, B.A.M., Noble, D., Foster, S.J., Webb, G., Haxell, M. (1990) The biosynthesis of a family of novel antiparasitic macrolides. Proceedings of the 6th International Symposium on the Genetics of Industrial Microorganisms. Strausbourg, France. Abstract A70. p.96. ISBN 2-87805-004-5.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbour, Laboratory Press. New York.

Schmidt, D.M.Z., Hubbard, B.K., and Gerlt, J.A. (2001) Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in *Bacillus subtilis* and *Escherichia coli* as L-Ala-D/L-Glu epimerases. *Biochemistry* **40**:15707-15715.

Schwecke, T., Aparicio, J.F., Molnár, I., König, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortés, J., Lester, J.B., Böhm, G.A., Staunton, J., and Leadlay, P.F. (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Nat. Acad. Sci. USA* **92**:7839-7843.

Shaw-Reid, C.A., Kelleher, N.L., Losey, H.C., Gehring, A.M., Berg, C., and Walsh, C.T. (1999) Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chem. Biol.* **6**:385-400.

Silakowski, B., Nordsiek, G., Kunze, B., Blöcker, H., and Müller, R (2001) Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantica* Sga15. *Chem. Biol.* **8**:59-69.

Singh, S.K., Gurusiddaiah, S., and Whalen, J.W. (1985) Treponemycin, a nitrile antibiotic active against *Treponema hyodysenteriae. Antimicrob. Agents Chemother.* **27**:239-245.

Staunton, J., and Wilkinson, B. (1997) Biosynthesis of erythromycin and rapamycin. *Chem. Rev.* **97**:2611-2629.

Swan, D.G., Rodriguez, A.M., Vilches, C., Mendez, C., and Salas, J.A. (1994) Characterization of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence. *Mol. Gen. Genet.* **242**:258-362.

Takeshita, S., Sato, M., Toba, M., Masahashi, W., and Hashimoto-Gotoh, T. (1987) High-copy number and low-copy number plasmid vectors for *lacZ* alpha-complementation and chloroamphenicol- or kanamycin-resistance selection. *Gene* **61**:63-74.

Thomas, I., Martin (née Rowe), C.J., Wilkinson, C.J., Staunton, J., and Leadlay, P.F. (2002) Skipping in a hybrid polyketide synthase: evidence for ACP to ACP chain transfer. *Chem. Biol.* **9**:781-787.

Tsuchiya, E., Yukawa, M., Miyakawa, T., Kimura, K.I., and Takahashi, H. (2001) Borrelidin inhibits a cyclin-dependent kinase (CDK), Cdc28/Cln2, of *Saccharomyces cerevisiae*. *J. Antibiot.* **54**:84-90.

Wakasugi, K., Slike, B.M., Hood, J., Otani, A., Ewalt, K.L., Friedlander, M., Cheresh, D.A., and Schimmel, P. (2002) A human aminoacyl-tRNA synthetase as a regulator of angiogenesis. *Proc. Nat. Acad. Sci. USA* **99**:173-177.

Wakabayashi, T., Kageyama, R., Naruse, N., Tsukahara, N., Funahashi, Y., Kitoh, K., and Watanabe, Y. (1997) Borrelidin is an angiogenesis inhibitor; disruption of angiogenic capilla vessels in a rat aorta matrix culture model. *J. Antibiot.* **50**:671-676.

Waldron, C., Matsushima, P., Rosteck, P.R., Broughton, M.C., Turner, J., Madduri, K., Crawford, K.P., Merlo, D.J. and Baltz, R.H. (2001) Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*. *Chem.Biol.* **8**:487-499.

Wilkinson, B., Foster, G., Rudd, B.A.M., Taylor, N.L., Blackaby, A.P., Sidebottom, P.J., Dawson, M.J., Buss, A.D., Gaisser, S., Böhm, I.U., Rowe, C.J., Cortés, J., Leadlay, P.F., and Staunton, J. (2000) Novel octaketide macrolides related to 6-deoxyerythronolide B provide

evidence for iterative operation of the erythromycin polyketide synthase. *Chem. Biol.* **7**:111-117.

Wu, N., Tsuji, S.Y., Cane, D.E., and Khosla, C. (2001) Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc.* **123**: 6465-6474.

Xue, Y.Q., Zhao, L.S., Liu, H.-W., and Sherman, D.H. (1998) A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *Proc. Nat. Acad. Sci. USA* **95**:12111-12116.

Xue, Y.Q., and Sherman, D.H. (2000) Alternative modular polyketide synthase expression controls macrolactone structure. *Nature* **403**:571-575.

Yamamoto, H., Maurer, K.H., Hutchinson, C.R. (1986) Transformation of *Streptomyces erythraeus*. *J. Antibiot.* **39**:1304-1313.